

Le, Emily

From: Le, Emily
Sent: Tuesday, August 03, 2004 9:45 AM
To: Delaval, Jan
Subject: RE: sequence search: 09/939537

-----Original Message-----

From: Delaval, Jan
Sent: Tuesday, August 03, 2004 5:58 AM
To: Le, Emily
Subject: RE: sequence search: 09/939537

I need to know the exact points of fusion for each request., example number 3: 1-394 of # 29 fusion to position 1 of # 32.

-----Original Message-----

From: Le, Emily
Sent: Monday, August 02, 2004 9:10 AM
To: Delaval, Jan
Subject: sequence search: 09/939537

Hi Jan,

Please provide a search of:

- ~~**[Le, Emily]** 1. Residues 1-395 of SEQ ID NO: 29~~
- ~~2. Residues 1-200 of SEQ ID NO: 31~~
- 3. Residues 1-394 of SEQ ID NO: 29 fused with SEQ ID NO: 32 **[Le, Emily]** , at position 1 of SEQ ID NO: 32
- 4. Residues 1-200 of SEQ ID NO: 31 fused with SEQ ID NO: 32 **[Le, Emily]** at position 1 of SEQ ID NO: 32
- 5. Residues 1-395 of SEQ ID NO: 29 fused with SEQ ID NO: 35 **[Le, Emily]** at position 1 of SEQ ID NO: 35
- 6. Residues 1-200 of SEQ ID NO: 31f used with SEQ ID NO: 35 **[Le, Emily]** at position 1 of SEQ ID NO: 35
- 7. Residues 1-394 of SEQ ID NO: 29 fused with SEQ ID NO: 32 **[Le, Emily]** at position 1 of SEQ ID NO: 32, fused with SEQ ID NO: 35 **[Le, Emily]** at position 1 of SEQ ID NO: 35
- 8. Residues 1-200 of SEQ ID NO: 31 fused with SEQ ID NO: 32 **[Le, Emily]** at position 1 of SEQ ID NO: 32, fused with SEQ ID NO: 35 **[Le, Emily]** at position 1 of SEQ ID NO: 35

for the above case. Please provide a printout of the first 100 hits. Thanks, Jan.

Emily Le
Mailbox, Remsen 3C18
Office, Remsen, 3C35
(571) 272-0903

THIS PAGE BLANK (USPTO)

: Entered [jdelaval 3-Aug-04 11:02]

SE01
MNRGVPFRHLLVQLALLPATOQNKVVLGKGDVTELTCTASQKSIQFHMKNNOIKILGNGSFLT
KQPSKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVOLVFGJLANS DTHLLOQOSLT
LTLESPPSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTWCTVLONOKKVEFKIDIVL
VYKKEGEVFEFPLAFTEVKELTGSGELMWOAEARASSSKWITFDLKNKEVSVKRVTOPKLOMGKCLPL
HLLFOALPOVAGSGNLTLALEAKTGKLEHDEVNLVVRATOLOKNLTCEVWGPTSPKMLSLKLENKEAK
VSKREKPVWVLNPEAGMOCCLSDSGQVLESNIKVLPTWSTPVIH

: Entered [jdelaval 3-Aug-04 11:05]

SE02
MNRGVPFRHLLVQLALLPATOQNKVVLGKGDVTELTCTASQKSIQFHMKNNOIKILGNGSFLT
KQPSKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVOLVFGJLANS DTHLLOQOSLT
LTLESPPSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTWCTVLONOKKVEFKIDIVL

; Entered [jdelaval 3-Aug-04 11:02]
SFO3
MRGVPFRHLLLVLOLALLPAAATGKNVVLGKKGDVTELTCTASQKKSIOFHMKNNSNOIKILGNOSFLT
KGPSKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVQLVFGLTANSDBTHLQGSFLT
LTLESPPGSSPSVOCSPRGKNIQGGKTLVSQLELQDSGTWCTVLOKQKVEFKIDIVLAQKASSI
VYKKEGEQVEFSFPLAFYVEKLTGSGELMWOAERASSSKSWITFDLKNKEVSVKRYTODPKLQMGKCLPL
HLTLFOALPOVYAGSGNLTLAEAKTGKLEHQBVLVVMRATOLQKNTCEVMGPTSPQLMLSLKENKEAK
VSKREKPVVNLNPEAGMOCILSDSGOVLLESNIKVLPTWSTPVEBKSCDKHTCPCPAPELGGPSVF
LEPPKDKOTLMTSRTEVTCVVVDVSHEDPEVKFNNYVDCGVENNAKTKPREEOYNSUTRYVSVLTVLHQ
DMLNGKEYCKCKVSNKALPAPIEKTISKAKGQPREPOVYTTLPSSRDELTKNQVSLTCLVKGFTPSDIAVEH
ESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSRMQQGNVFPSCSVMEHALHNHYTQKSLSLSPGLQDE
TCAEAQDGLDGLMTTDP1

; Entered [jdelaval 3-Aug-04 11:05]
SBO4
MRGVPFRHLLLVLOLALLPAAATGKNVVLGKKGDVTELTCTASQKKSIOFHMKNNSNOIKILGNOSFLT
KGPSKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVQLVFGLTANSDBTHLQGSFLT
LTLESPPGSSPSVOCSPRGKNIQGGKTLVSQLELQDSGTWCTVLOKQKVEFKIDIVPCPAPEPKSC
DKHTCPELLGGPSVFLFPPKDKOTLMTSRTEVTCVVVDVSHEDPEVKFNNYVDCGVENNAKTKPREEO
YNSUTRYVSVLTVLHQDMLNGKEYCKVSNKALPAPIEKTISKAKGQPREPOVYTTLPSSRDELTKNQVSL
TCLVKGFTPSDIAVEHESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSRMQQGNVFPSCSVMEHALHNH
YTQKSLSLSPGLQDETCAEAQDGLDGLMTTDP1

! Entered [jdelaval 3-Aug-04 11:02]
SPO5
MRGVPRFRLLVLOLALLPATQGNKVVLGKGDVVELTCTASQKSIQFHMKNNOIKILGNQGSFLT
KGPSKLNDRADSRSLMDQGNPLIINKLIEDSDTYICEVEDQKEVQLVFGLTANSDTHLLOQSFLT
LTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNKQKVEFKIDI VPLAFOKASSI
VYKSGEQVEFSFPLAFTEKLTGSGELMWOAEARASSSKSWITFDLKNKEVSKRVTOPLKQMGKCLPL
HLLPOALPOVAGSGNLTLEAKTGKLEHOENVLYVMRATOLQKNTCEVMGPTSPKMLSLKLEKCAK
VSKREKPVVNLNBPAGMOCILSDSGOVILLESNIKVLPWTSTPVHPRASALPAPPTGSALPDPQTASALP
DPPASALPALAVISFLGLGIGVACVILARTRI

3

! Entered [jdelaval 3-Aug-04 11:05]
SPO6
MRGVPRFRLLVLOLALLPATQGNKVVLGKGDVVELTCTASQKSIQFHMKNNOIKILGNQGSFLT
KGPSKLNDRADSRSLMDQGNPLIINKLIEDSDTYICEVEDQKEVQLVFGLTANSDTHLLOQSFLT
LTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNKQKVEFKIDI VPRASALPAP
TGSALPDPQTASALPDPASALPALAVISFLGLGIGVACVILARTRI

29 \$

; Entered [jdelaval 3-Aug-04 11:02]
SEO7
MNRGVPFRHLLVLOLALLPAATOGNKVVLGKKGDIVELTCTASOKKSIOPHMKNSNOIKILGNGSFLT
KGPSTLNRADSRSRSLMDOGNPPLIKNLKIEDSDTYICEVEDQKEVQLLVFGLTANSDTHLLOGOST
LTLESPPGSSPSVQCSFRGKNIOGGKTLVSQLELODSGTWCTVLQNKKEVEFKIDI VLAFOQASIT
VYKKEGEQVESFPPLAFVEKLTGSELWQARASSSSKSWITFDLKNKEVSVKRVTODPKLOMGKCLPL
HLTLPOALPOVAGSGNLTLALEAKTGKLEHQBENVLVVWRATOLQKNLTCVWGSPKJMLSLKENKEAK
VSKREKPVVNLNPEAGMOCLLSDSGOVLLESNIKVLPTWSTFVPCPAPRPSKCDKHTCPELGGPSVF
LPPKPKDITLMI SRPEVTCVVVDVSHEDPEVKFNMYVDGVEVHNAKTKRPREOYNSTYRVVSVLTVLHQ
DMLNCKEYCKVSNKALPAPIEKTISKAKGQPREPOVYTTLPBSRDELITKNQUSITGLVKGFTYPSDIAVEN
ESNGQPENNYKTTPVLDSDGSFFLYSKLTVDKSRMGOGNVFSCSVMHEALHNHTYQKSLSPGLQDDE
TCAEAQDDELGLWTTDPI

; Entered [jdelaval 3-Aug-04 11:05]
SEO8
MNRGVPFRHLLVLOLALLPAATOGNKVVLGKKGDIVELTCTASOKKSIOPHMKNSNOIKILGNGSFLT
KGPSTLNRADSRSRSLMDOGNPPLIKNLKIEDSDTYICEVEDQKEVQLLVFGLTANSDTHLLOGOST
LTLESPPGSSPSVQCSFRGKNIOGGKTLVSQLELODSGTWCTVLQNKKEVEFKIDI VPCPAPRPSK
DKHTCPELGGPSVFLPPKPKDITLMI SRPEVTCVVVDVSHEDPEVKFNMYVDGVEVHNAKTKPREO
YNSTYRVVSVLTVLHQDMLNCKEYCKVSNKALPAPIEKTISKAKGQPREPOVYTTLPBSRDELITKNQUSL
TCLVKGFTYPSDIAVENESNGQPENNYKTTPVLDSDGSFFLYSKLTVDKSRMGOGNVFSCSVMHEALHNH
YTOKSLSPGLQDDETCABAQDDELGLWTTDPPRASALPAPPTGSALPDPTASALPDPPASALPAA
LAVISFLGLGLGAVACVLAATRI

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using ew model

Run on: August 3, 2004, 13:00:24 ; Search time 14.8924 Seconds
(without alignments)
4185.504 Million cell updates/sec

Title: SEQ3
Perfect score: 3414
Sequence: 1 MNRGVPRHLLVLQLALLP.....DETCAADQGEJLMTTDP 648

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%
Listing first 125 summaries

Database : PIR 78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2024	59.3	458	1 RWHUT4	T-cell surface gly
2	1876	55.0	432	1 RWC2T4	T-cell surface gly
3	1738	50.9	432	1 RMCQT4	T-cell surface gly
4	1266	37.1	330	1 GHNU	Ig gamma-1 chain C
5	1254	36.7	255	4 S31866	Ig gamma-1 chain C
6	1252.5	36.7	374	2 S69339	Ig heavy chain V r
7	1218	35.7	234	2 P10207	Ig gamma-1 chain C
8	1171	34.3	377	2 A23511	Ig gamma-3 chain C
9	1169	34.2	377	2 A60764	Ig gamma-3 chain C
10	1151	33.7	289	1 G3HWT1	Ig gamma-3 heavy C
11	1150	33.7	326	1 G2HU	Ig gamma-2 chain C
12	1149	33.7	459	2 A46254	CD4 precursor - ra
13	1141	33.4	327	1 G4HU	Ig gamma-4 chain C
14	1105	32.4	432	2 S30193	T-cell surface gly
15	999	29.3	457	2 A27449	T-cell surface gly
16	993	29.1	457	1 RMMST4	T-cell surface gly
17	947.5	27.8	398	1 G3MSM	Ig gamma-3 chain C
18	940.5	27.5	393	1 G3MSM	Ig gamma-3 chain C
19	922	27.0	323	1 GMRB	Ig gamma-1 chain C
20	915.5	26.8	399	1 G2MSAM	Ig gamma chain C r
21	911.5	26.7	328	2 I47160	Ig gamma-2a chain
22	911.5	26.7	328	2 I47159	Ig gamma-2b chain
23	906.5	26.6	329	2 I47162	Ig gamma-2a chain
24	906.5	26.6	329	1 G2GP	Ig gamma-2 chain C
25	891.5	26.1	328	2 I47158	Ig gamma-1 chain C
26	884.5	25.9	405	1 G2MSBM	Ig gamma-2b chain
27	884.5	25.9	470	2 S22080	Ig heavy chain pre
28	883.5	25.9	328	2 I47161	Ig gamma-3 chain C
29	861	25.2	472	2 S31459	Ig gamma-1 chain -

30	851	24.9	308	2 C30554	Ig heavy chain C r
31	849.5	24.9	329	1 G3MSC	Ig gamma-3 chain C
32	844	24.7	333	2 PS0018	Ig gamma-2b chain
33	839.5	24.6	444	2 PC4436	monoclonal antibody
34	836.5	24.5	469	2 S37483	Ig gamma-2a chain
35	828	24.3	326	2 PS0017	Ig gamma-1 chain C
36	826.5	24.2	324	1 G1MS	Ig gamma-1 chain C
37	817.5	23.9	329	2 S00847	Ig gamma-2c chain
38	816	23.9	335	1 G2MSAB	Ig gamma-2a chain
39	814.5	23.9	330	1 G2MSA	Ig gamma-2a chain
40	812	23.8	474	1 G2MS11	Ig gamma-2b chain
41	811	23.8	446	2 S40295	Ig gamma-2a chain
42	795	23.3	322	2 PS0019	Ig gamma-2a chain
43	782.5	22.9	475	2 S01321	Ig gamma-2b chain
44	774	22.7	327	2 S06611	Ig gamma-2 chain C
45	702	20.6	180	2 I46732	Ig gamma heavy cha
46	574.5	16.8	218	2 A36040	Ig heavy chain V-I
47	572.5	16.8	249	2 S69340	Ig heavy chain VH1
48	566	16.6	152	2 S14236	Ig gamma-1 chain C
49	394.5	11.6	572	2 B46529	Ig Y heavy chain (
50	390.5	11.4	627	2 S14683	Ig mu chain precu
51	389	11.4	549	2 S04845	Ig heavy chain pre
52	375.5	11.0	476	1 MHMSM	Ig mu chain C regi
53	371.5	10.9	592	2 S25705	Ig mu chain - shee
54	370	10.8	548	2 S38864	Ig epsilon chain C
55	364	10.7	455	1 MHMS	Ig mu chain C regi
56	364	10.7	455	1 A24976	Ig mu chain C regi
57	363.5	10.6	388	1 EHMS	Ig epsilon chain C
58	360	10.5	474	2 S15590	Ig heavy chain - h
59	359.5	10.5	343	2 S25644	Ig mu chain C regi
60	359.5	10.5	448	2 S03186	Ig heavy chain C r
61	359.5	10.5	454	1 MHY	Ig mu chain C regi
62	359	10.5	428	1 EHTU	Ig epsilon chain C
63	358.5	10.5	453	2 S37768	Ig mu chain C regi
64	358.5	10.5	479	1 MHRBM	Ig mu chain C regi
65	358	10.5	504	2 S00390	Ig gamma chain (cl
66	357	10.5	71	2 I60082	CD4 receptor - hum
67	355.5	10.4	429	1 EHRT	Ig epsilon chain C
68	353.5	10.4	391	1 MHRBT	Ig mu heavy chain
69	353	10.3	426	2 I36948	Ig epsilon-chain -
70	352.5	10.3	473	1 MHRUM	Ig mu chain C regi
71	349	10.2	458	1 MHRB	Ig mu chain C regi
72	348	10.2	452	1 MHU	Ig mu chain C regi
73	348	10.2	457	2 S03961	Ig mu chain C regi
74	345	10.1	423	1 EHMS	Ig epsilon chain C
75	340.5	10.0	577	2 I50731	Ig heavy chain - n
76	340	10.0	450	1 MHG	Ig mu chain C regi
77	333.5	9.8	453	2 C31933	Ig gamma-3 chain C
78	331.5	9.7	112	2 B30503	Ig gamma-2a chain
79	330.5	9.7	684	2 S60266	Ig gamma-2a chain
80	313.5	9.2	433	2 S31436	Ig gamma-2a chain
81	305.5	8.9	99	2 S21461	Ig gamma-2a chain
82	305	8.9	299	1 AHRB	Ig alpha chain C r
83	304.5	8.9	357	2 S09269	Ig alpha chain C r
84	302	8.8	338	2 S09276	Ig alpha chain C r
85	297	8.7	438	1 HVRKC2	Ig mu chain C regi
86	296	8.7	342	2 I47175	Ig alpha chain C r
87	295.5	8.7	348	2 S09270	Ig alpha chain C r
88	295.5	8.6	585	2 A46507	Ig alpha chain - c
89	295	8.6	461	1 HVRKC0	Ig mu chain C regi
90	293.5	8.6	357	2 S09265	Ig alpha chain C r
91	293.5	8.6	438	1 HVRKCS	Ig mu chain C regi
92	292.5	8.6	347	2 S09274	Ig alpha chain C r
93	291	8.5	367	2 MHCH	Ig mu chain C regi
94	285.5	8.4	343	2 S09272	Ig alpha chain C r
95	284	8.4	244	2 S12328	Ig heavy chain C r
96	282.5	8.3	340	2 B23360	Ig alpha-2 chain C
97	281.5	8.2	393	2 HVRKCL	Ig mu chain C regi
98	280.5	8.2	99	2 S21462	T-cell surface gly
99	280.5	8.2	352	2 S09266	Ig alpha chain C r
100	280.5	8.2	353	1 ALHU	Ig alpha-1 chain C
101	279	8.2	339	2 S09264	Ig alpha chain C r
102	279	8.2	568	2 A45804	Ig mu chain C regi

```
103 278.5 8.2 474 2 150830 Ig mu chain - Lepi
104 278 8.1 352 2 505500 Ig alpha-1 chain C
105 277.5 8.1 357 2 S09267 Ig alpha chain C r
106 276.5 8.1 340 2 156230 Ig alpha-2 chain -
107 275.5 8.1 88 2 A30503 Ig gamma-2b chain
108 275 8.1 568 2 A34891 Ig heavy chain pre
109 270.5 7.9 358 2 S09268 Ig alpha chain C r
110 269 7.9 360 2 S09271 Ig alpha chain C r
111 268 7.9 573 2 S12838 Ig mu chain precu
112 267 7.8 370 1 HVK8C5 Ig mu chain C regi
113 266.5 7.7 348 2 S09273 Ig alpha chain C r
114 263 7.6 342 1 A2HU Ig alpha-2 chain C
115 259.5 7.6 340 2 A45966 Ig alpha chain C r
116 257 7.5 111 2 S43148 Ig upsilon chain -
117 257 7.5 454 2 A46532 Ig mu chain C regi
118 249 7.3 335 2 S09275 Ig alpha chain C r
119 248.5 7.3 344 1 AAMS Ig alpha chain C r
120 245 7.2 1005 T18537 Ig heavy chain C r
121 243.5 7.1 220 2 C22360 Ig alpha-2 chain C
122 240.5 7.0 580 2 A46538 Ig heavy chain, se
123 211 6.2 448 2 A46533 Ig heavy chain C r
124 209 6.1 115 2 168731 IGE chain C4 regio
125 206 6.0 402 2 S20002 Ig heavy chain, se
```

ALIGNMENTS

RESULT 1

RNMUT4 T-cell surface glycoprotein CD4 precursor [validated] - human

N/Alternate names: T-cell surface antigen T4/Lew 3

C/Species: Homo sapiens (man)

C/Date: 28-May-1996 #sequence revision 31-Dec-1998 #text change 20-Apr-2001

C/Accession: A080872; A32722; A34194; A53287; I54176; I54297; A02109; A30039

R/Madon, P.J.; Littman, D.R.; Godfrey, W.; Madon, D.E.; Chess, L.; Axel, R.

Cell 42, 93-104, 1985

A/Title: The isolation and nucleotide sequence of a cDNA encoding the T cell surface pro

A/Reference number: A90872; MUID:85254948; PMID:2990730

A/Accession: A90872

A/Molecule type: mRNA

A/Residues: 1-25, 'N', 27-458 <MAD>

A/Experimental source: clone PT48

R/Littman, D.R.; Madon, P.J.; Axel, R.

Cell 55, 541, 1988

A/Title: Corrected CD4 sequence.

A/Reference number: A90907; MUID:89028665; PMID:3263213

A/Contents: annotation; revision to residue 26

R/Camerini, D.; Seed, B.

Cell 60, 747-754, 1990

A/Title: A CD4 domain important for HIV-mediated syncytium formation lies outside the vi

A/Accession: A32722

A/Status: nucleic acid sequence not shown; not compared with conceptual translation

A/Molecule type: mRNA

A/Residues: 26-426, 428-458 <CAM>

R/Carr, S.A.; Hemling, M.E.; Foltens-Maaserman, G.; Sweet, R.W.; Annunula, K.; Barr, J.R.;

J. Biol. Chem. 264, 21286-21295, 1989

A/Title: Protein and carbohydrate structural analysis of a recombinant soluble CD4 recep

A/Contents: disulfide bonds; carbohydrate-binding sites

A/Accession: A34194

A/Molecule type: protein

A/Residues: 26-394 <CAR>

R/Lederman, S.; Demattino, J.A.; Daugherty, B.L.; Foeldvari, I.; Yellin, M.J.; Cleary, A

Mol. Immunol. 28, 1171-1181, 1991

A/Title: A single amino acid substitution in a common African allele of the CD4 molecule

A/Reference number: A53287; MUID:92072595; PMID:1961196

A/Accession: A53287

A/Status: not compared with conceptual translation

A/Molecule type: mRNA

A/Residues: 250-264, 'W', 266-280 <LED>

A/Note: sequence extracted from NCBI backbone (NCBIP:68249)

R/Edwards, M.C.; Gibbs, R.A.

Genomics 14, 590-597, 1992

A/Title: A human dimorphism resulting from loss of an Alu.

A/Reference number: I54176; MUID:93052387; PMID:1330888

A/Accession: I54176

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-72 <RES>

R/Hodge, T.W.; Sasse, D.R.; McDougal, J.S.

Hum. Immunol. 30, 99-104, 1991

A/Title: Humans with OKT4-epitope deficiency have a single nucleotide base change in the

A/Reference number: I54297; MUID:91216786; PMID:11708753

A/Accession: I54297

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-264, 'W', 266-458 <RES2>

A/Cross-references: GB:M35160; NID:g179143; PIDN:AAA16069.1; PID:g179144

C/Comment: Macrophage tropic strains of HIV-1 bind to a complex of chemokine (C-C) recep

C/Genetics:

A/Genes: GDB:CD4

A/Cross-references: GDB:119767; OMIM:186940

A/Map position: 12pter-12p12

A/Introns: 16/3

C/Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology

C/Keywords: AIDS; duplication; glycoprotein; T-cell; transmembrane protein

F/1-25/Domain: signal sequence #status predicted <SIG>

F/26-458/Product: T-cell surface glycoprotein CD4 #status experimental <MAT>

F/34-111/Domain: immunoglobulin homology <IM1>

F/136-186/Domain: immunoglobulin homology #status atypical <IM2>

F/216-299/Domain: immunoglobulin homology <IM3>

F/321-372/Domain: immunoglobulin homology <IM4>

F/397-420/Domain: transmembrane #status predicted <TM>

F/421-458/Domain: intracellular #status predicted <INT>

F/41-109,155-184,328-370/Disulfide bonds: #status experimental

F/236,325/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 59.3%; Score 2024; DB 1; Length 458;

Best Local Similarity 99.2%; Pred. No. 1,7e-112;

Matches 393; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLLVQLALPAATQGNKRVIGKKDVELTCTASQKKSIOFHMNSNOIK 60

Db 1 MNRGVPFRHLLVQLALPAATQGNKRVIGKKDVELTCTASQKKSIOFHMNSNOIK 60

QY 61 ILNQGSFLLTKGSPSKNDRASSRSIMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVOL 120

Db 61 ILNQGSFLLTKGSPSKNDRASSRSIMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVOL 120

QY 121 LVFGLTANSDTHLLQGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG 180

Db 121 LVFGLTANSDTHLLQGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG 180

QY 181 TWTCVTLQNGKVEFKIDIVIAFQKASSIVYKKEGQVEFSPPLAFTVEKLTGSGELMW 240

Db 181 TWTCVTLQNGKVEFKIDIVIAFQKASSIVYKKEGQVEFSPPLAFTVEKLTGSGELMW 240

QY 241 QARRASSKSWITFDLKNKEVSIVKRTQDPKLMGKGLPHLTLPLQALPOYASSGNITLA 300

Db 241 QARRASSKSWITFDLKNKEVSIVKRTQDPKLMGKGLPHLTLPLQALPOYASSGNITLA 300

QY 301 LEAKTGKLEHVEVLVVMRATQKNLTCEWGPSPKLMSTLKENEAIVSKREKPVWV 360

Db 301 LEAKTGKLEHVEVLVVMRATQKNLTCEWGPSPKLMSTLKENEAIVSKREKPVWV 360

QY 361 LNDEAGMOCCLLSDSGQVLLSNIKVLPWTSTVEVP 396

Db 361 LNDEAGMOCCLLSDSGQVLLSNIKVLPWTSTVEVP 396

RESULT 2

RWC274 T-cell surface glycoprotein CD4 - chimpanzee

QY 361 LNDEAGMOCCLLSDSGQVLLSNIKVLPWTSTVEVP 396

Db 361 LNDEAGMOCCLLSDSGQVLLSNIKVLPWTSTVEVP 396

N/Alternate names: T-cell surface antigen T4/Len 3
C/Species: Pan troglodytes (chimpanzee)
C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
C/Accession: B32722; A46534
R/Camerini, D.; Seed, B.
Cell 60, 747-754, 1990
A/Title: A CD4 domain important for HIV-mediated syncytium formation lies outside the vi
A/Reference number: A32722; MUID:90182664; PMID:2107024
A/Accession: B32722
A/Molecule type: mRNA
A/Residues: 1-432 <CAM>
A/Cross-references: GB:M31135
R/Fomsgaard, A.; Hirsch, V.M.; Johnson, P.R.
Eur. J. Immunol. 22, 2973-2981, 1992
A/Title: Cloning and sequences of primate CD4 molecules: diversity of the cellular recep
A/Reference number: A46534; MUID:93049640; PMID:1425921
A/Accession: A46534
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 3-339 <ROM>
A/Note: sequence extracted from NCBI backbone (NCBP:118332)
C/Comment: This protein is expressed on most thymocytes, on a subset of mature T-cells
C/Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
C/Keywords: duplication; glycoprotein; T-cell; transmembrane protein
F:1-432/Product: T-cell surface glycoprotein CD4 #status predicted <MAT>
F:1-371/Domain: extracellular #status predicted <EXT>
F:9-86/Domain: immunoglobulin homology <IM1>
F:111-161/Domain: immunoglobulin homology #status atypical <IM2>
F:191-274/Domain: immunoglobulin homology <IM3>
F:296-347/Domain: immunoglobulin homology <IM4>
F:372-395/Domain: transmembrane #status predicted <TM>
F:396-432/Domain: intracellular #status predicted <INT>
F:16-84,130-159,303-345/Disulfide bonds: #status predicted
F:271,300/Binding site: carbonylate (Asn) (covalent) #status predicted

Query Match 50.0%; Score 1876; DB 1; Length 432;
Best Local Similarity 98.4%; Pred. No. 9, 1e-104;
Matches 364; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 27 KVLGKGGDTVELTCTASQKKSIOFHMKNSNQIKILGNQGSFLTKGPSKLNDRADSRSL 86
DB 2 KVLGKGGDTVELTCTASQKKSIOFHMKNSNQIKILGNQGSFLTKGPSKLNDRADSRSL 61
QY 87 WQGNFPLIKLKLTEDSDTYICEVEDQKEVQLLVFGLTANSDFHLQGGSLTTLTLESP 146
DB 62 WQGNFPLIKLKLTEDSDTYICEVEDQKEVQLLVFGLTANSDFHLQGGSLTTLTLESP 121
QY 147 PSSSPVQCRSPRGNIOGGKTLVSQLELDGSGTWCTVLQNKVEFKIDIVLAFOK 206
DB 122 PSSSPVQCRSPRGNIOGGKTLVSQLELDGSGTWCTVLQNKVEFKIDIVLAFOK 181
QY 207 ASSIYKKEGQEVESFPLAFTVEKLTGSGELMWQAEKRASSKSWITFDLNKKEVSVKRV 266
DB 182 ASSIYKKEGQEVESFPLAFTVEKLTGSGELMWQAEKRASSKSWITFDLNKKEVSVKRV 241
QY 267 TCEVNGPSPKLMSTLKENKAKVSKREKPVVNLNPEAGMOCCLSDSGOVLLESNIKV 386
DB 302 TCEVNGPSPKLMSTLKENKAKVSKREKAVVNLNPEAGMOCCLSDSGOVLLESNIKV 361
QY 387 LPTWSTPVP 396
DB 362 LPTWSTPVP 371

RESULT 3
R/MQ74
T-cell surface glycoprotein CD4 - rhesus macaque
N/Alternate names: T-cell surface antigen T4/Len 3
C/Species: Macaca mulatta (rhesus macaque)

C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
C/Accession: C32722
R/Camerini, D.; Seed, B.
Cell 60, 747-754, 1990
A/Title: A CD4 domain important for HIV-mediated syncytium formation lies outside the vi
A/Reference number: A32722; MUID:90182664; PMID:2107024
A/Accession: C32722
A/Molecule type: mRNA
A/Residues: 1-432 <CAM>
A/Cross-references: GB:M31134
C/Comment: This protein is expressed on most thymocytes, on a subset of mature T-cells
C/Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
C/Keywords: duplication; glycoprotein; T-cell; transmembrane protein
F:1-432/Product: T-cell surface glycoprotein CD4 #status predicted <MAT>
F:1-371/Domain: extracellular #status predicted <EXT>
F:9-86/Domain: immunoglobulin homology <IM1>
F:111-161/Domain: immunoglobulin homology #status atypical <IM2>
F:180-293/Domain: immunoglobulin homology <IM3>
F:296-347/Domain: immunoglobulin homology <IM4>
F:372-395/Domain: transmembrane #status predicted <TM>
F:396-432/Domain: intracellular #status predicted <INT>
F:16-84,130-159,303-345/Disulfide bonds: #status predicted
F:271,300/Binding site: carbonylate (Asn) (covalent) #status predicted

Query Match 50.9%; Score 1738; DB 1; Length 432;
Best Local Similarity 90.5%; Pred. No. 1, 3e-95;
Matches 335; Conservative 17; Mismatches 18; Indels 0; Gaps 0;

QY 27 KVLGKGGDTVELTCTASQKKSIOFHMKNSNQIKILGNQGSFLTKGPSKLNDRADSRSL 86
DB 2 KVLGKGGDTVELTCTASQKKSIOFHMKNSNQIKILGNQGSFLTKGPSKLNDRADSRSL 61
QY 87 WQGNFPLIKLKLTEDSDTYICEVEDQKEVQLLVFGLTANSDFHLQGGSLTTLTLESP 146
DB 62 WQGNFPLIKLKLTEDSDTYICEVEDQKEVQLLVFGLTANSDFHLQGGSLTTLTLESP 121
QY 147 PSSSPVQCRSPRGNIOGGKTLVSQLELDGSGTWCTVLQNKVEFKIDIVLAFOK 206
DB 122 PSSSPVQCRSPRGNIOGGKTLVSQLELDGSGTWCTVLQNKVEFKIDIVLAFOK 181
QY 207 ASSIYKKEGQEVESFPLAFTVEKLTGSGELMWQAEKRASSKSWITFDLNKKEVSVKRV 266
DB 182 ASSIYKKEGQEVESFPLAFTVEKLTGSGELMWQAEKRASSKSWITFDLNKKEVSVKRV 241
QY 267 TCEVNGPSPKLMSTLKENKAKVSKREKPVVNLNPEAGMOCCLSDSGOVLLESNIKV 386
DB 302 TCEVNGPSPKLMSTLKENKATVSKQAKAVVNLNPEAGMOCCLSDSGOVLLESNIKV 361
QY 387 LPTWSTPVP 396
DB 362 VPTWPTVP 371

RESULT 4
GHU
Ig gamma-1 chain C region - human
C/Species: Homo sapiens (man)
C/Date: 31-Jan-1981 #sequence_revision 18-Aug-1982 #text_change 16-Jul-1999
C/Accession: A93433; S36861; S33887; B90563; A90564; B91668; A91723; A02146
R/Ellison, J.W.; Berson, B.J.; Hood, L.E.
Nucleic Acids Res. 10, 4071-4079, 1982
A/Title: The nucleotide sequence of a human immunoglobulin C-gamma1 gene.
A/Reference number: A93433; MUID:82274238; PMID:6287432
A/Accession: A93433
A/Molecule type: DNA
A/Residues: 1-330 <ELL>
A/Cross-references: EMBL:217370
A/Note: this sequence has the Gln(17) allelic marker, 97-Lys, and the Gln(1) markers,
A/Note: Lys-330 is removed after translation

R;Harris, L.J.
submitted to the EMBL Data Library, October 1992
A:Reference number: S33904
A:Accession: S36861
A:Molecule type: DNA
A:Residues: 2-330 <HAR>
A:Cross-references: EMBL:Z17370
R;Yakubashvili, N.; Ueda, S.; Odate, M.; Nikaido, T.; Nakai, S.; Honjo, T.
Cell 29, 671-679, 1982
A:Title: Structure of human immunoglobulin gamma genes: implications for evolution of a
A:Reference number: S33887; MUID:83001943; PMID:6811139
A:Accession: S33887
A:Molecule type: DNA
A:Residues: 88-113/235-330 <TAK>
A:Cross-references: EMBL:Z17370
R;Cunningham, B.A.; Rutishauser, U.; Gall, W.E.; Gottlieb, P.D.; Maxdall, M.J.; Edelman,
Biochemistry 9, 3161-3170, 1970
A:Title: The covalent structure of a human gammaG-immunoglobulin. VII. Amino acid sequen
A:Reference number: A90563; MUID:71064024; PMID:5489771
A:Contents: myeloma protein Eu
A:Accession: B90563
A:Molecule type: protein
A:Residues: 1-96, 'R', 98-135 <GUN>
A:Note: this sequence has the G1m(3) marker, 97-Arg
R;Rutishauser, U.; Cunningham, B.A.; Bennett, C.; Konigsberg, W.H.; Edelman, G.M.
Biochemistry 9, 3171-3181, 1970
A:Title: The covalent structure of a human gammaG-immunoglobulin. VIII. Amino acid sequen
A:Reference number: A90564; MUID:71064025; PMID:5530842
A:Contents: Eu
A:Accession: A90564
A:Molecule type: protein
A:Residues: 136-154, 'Q', 156-165, 'Q', 167-176, 'Q', 178-194, 'N', 196-197, 'D', 199-238, 'E', 240,
A:Note: this sequence has the G1m(non-1) markers, 239-Glu and 241-Met
R;Ponstingl, H.; Hilschmann, N.
Hoppe-Seyler's Z. Physiol. Chem. 357, 1571-1604, 1976
A:Title: Die Primerstruktur eines monoklonalen IgG1-Immunglobulins (Myelomprotein Nie),
Igen Primerstruktur.
A:Reference number: A91668; MUID:77070269; PMID:826475
A:Contents: myeloma protein Nie
A:Accession: B91668
A:Molecule type: protein
A:Residues: 1-34, 'Q', 36-96, 'K', 98-115, 'Q', 117-197, 'D', 199-238, 'D', 240, 'L', 242-268, 'E', 27
A:Note: this sequence has the G1m(17) and G1m(1) markers
R;Schmidt, W.E.; Jung, H.D.; Palm, W.; Hilschmann, N.
Hoppe-Seyler's Z. Physiol. Chem. 364, 713-747, 1983
A:Title: Die Primerstruktur des kristallisierten monoklonalen Immunglobulins IgG1 KOL
A:Reference number: A91723; MUID:83289131; PMID:6884994
A:Contents: myeloma protein KOL; disulfide bonds
A:Accession: A91723
A:Molecule type: protein
A:Residues: 1-96, 'R', 98-197, 'D', 199-238, 'E', 240, 'M', 242-266, 'D', 268-271, 'D', 273-330 <SCH
A:Note: this sequence has the G1m(3) and G1m(non-1) markers
R;Gall, W.E.; Edelman, G.M.
Biochemistry 9, 3188-3196, 1970
A:Title: The covalent structure of a human gammaG-immunoglobulin. X. Intrachain disulfide
A:Reference number: A90565; MUID:71064027; PMID:4923144
A:Contents: annotation; disulfide bonds
R;Dirker, L.; Schwarz, J.; Reichel, W.; Hilschmann, N.
Hoppe-Seyler's Z. Physiol. Chem. 357, 1515-1540, 1976
A:Title: Rule of antibody structure. The primary structure of monoclonal IgG1 immunoglob
entworfene cleavage products, and the disulfide bridges.
A:Reference number: A91667; MUID:77070267; PMID:1002129
A:Contents: annotation; disulfide bonds
C:Genetics:
A:Gene: GDB:IGHG1
A:Cross-references: GDB:120085, OMTM:147100
A:Map position: 14q32.33-14q32.33
A:introns: 99/1; 114/1; 224/1
C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa
chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: duplication; glycoprotein; heterotetramer; immunoglobulin
F;20-85/Domain: immunoglobulin homology <IMI>

F;137-206/Domain: immunoglobulin homology <IM2>
F;243-310/Domain: immunoglobulin homology <IM3>
F;27-83,144-204,250-308/Disulfide bonds: #status experimental
F;103/Disulfide bonds: interchain (to light chain) #status experimental
F;109,112/Disulfide bonds: interchain (to heavy chain) #status experimental
F;180/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 37.1%; Score 1266; DB 1; Length 330;
Best Local Similarity 96.7%; Pred. No. 8e-68;
Matches 236; Conservative 0; Mismatches 4; Indels 4; Gaps 1;

QY 382 SNIKVLPFTWSTPVPKSCDKTHITCPCPAPELLGGSVFLFPPPKDTLMISRTPEVTCV 441
|||
DB 90 SNTKCV---DKVPEPKSCDKTHITCPCPAPELLGGSVFLFPPPKDTLMISRTPEVTCV 145
QY 442 VVDVSHDEPVKFMWYDGVVHNAKTRPEEQVNSYRVSVLTVLHODMLNGEKYCK 501
|||||
DB 146 VVDVSHDEPVKFMWYDGVVHNAKTRPEEQVNSYRVSVLTVLHODMLNGEKYCK 205
QY 502 VSNKALPAPLEKTISSAKGQPREPQVYTLPPSRDELTKNOVSLTCLVKGFYPSDIAVEME 561
|||
DB 206 VSNKALPAPLEKTISSAKGQPREPQVYTLPPSRDELTKNOVSLTCLVKGFYPSDIAVEME 265
QY 562 SNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSRWQGNVFCGVHEDALHNHYTQKSLS 621
|||
DB 266 SNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSRWQGNVFCGVHEDALHNHYTQKSLS 325
QY 622 LSPG 625
|||
DB 326 LSPG 329

RESULT 5
S31866
Ig gamma-1 chain C region - synthetic
C:Species: synthetic
A:Note: Homo sapiens (man) gene engineered and expressed in Escherichia coli
C:Date: 06-Jan-1995 #sequence_revision 17-Mar-1997 #text_change 19-May-2000
C:Accession: S31866
R;Filipula, D.
submitted to the EMBL Data Library, February 1993
A:Description: Screening method for protein-protein interactions of cloned gene products.
A:Reference number: S31866
A:Accession: S31866
A:Molecule type: mRNA
A:Residues: 1-255 <FIL>
A:Cross-references: EMBL:X70421; NID:g33068; PIDN:CAA49866.1; PID:g33069
C:Keywords: immunoglobulin
F;1-22/Region: Escherichia coli outer membrane protein A precursor
F;23-255/Region: human Ig gamma-1 chain C region

Query Match 36.7%; Score 1254; DB 4; Length 255;
Best Local Similarity 99.6%; Pred. No. 3e-67; Indels 0; Gaps 0;
Matches 231; Conservative 0; Mismatches 1;

QY 394 VEPKSCDKTHITCPCPAPELLGGSVFLFPPPKDTLMISRTPEVTCVVDVSHDEPVK 453
|||
DB 23 VEPKSCDKTHITCPCPAPELLGGSVFLFPPPKDTLMISRTPEVTCVVDVSHDEPVK 82
QY 454 FNNYVDGVVHNAKTRPEEQVNSYRVSVLTVLHODMLNGEKYCKKVSNAKLPAPLEK 513
|||||
DB 83 FNNYVDGVVHNAKTRPEEQVNSYRVSVLTVLHODMLNGEKYCKKVSNAKLPAPLEK 142
QY 514 TISKAKGQPREPQVYTLPPSRDELTKNOVSLTCLVKGFYPSDIAVEMSNQPENNYKTT 573
|||||
DB 143 TISKAKGQPREPQVYTLPPSRDELTKNOVSLTCLVKGFYPSDIAVEMSNQPENNYKTT 202
QY 574 PVLVDSDGSFFLYSKLTVDKSRWQGNVFCGVHEDALHNHYTQKSLSLSPG 625
|||||
DB 203 PVLVDSDGSFFLYSKLTVDKSRWQGNVFCGVHEDALHNHYTQKSLSLSPG 254

RESULT 6

```

569339
19 gamma chain V region precursor - human
C|Species: Homo sapiens (man)
C|Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 01-Dec-2000
C|Accession: S69339; 872664
R|Khamilich, A.A.; Aucturiarier, P.; Preud'homme, J.L.; Cogne, M.
Eur. J. Biochem. 229, 54-60, 1995
A|Title: Structure of abnormal heavy chains in human heavy-chain-deposition disease.
A|Reference number: S69339; PMID:95262687; PMID:7744049
A|Accession: S69339
A|Status: preliminary
A|Molecule type: mRNA
A|Residues: 1-374 <KHA>
A|Cross-references: EMBL:X81695
R|Khamilich, A.A.
submitted to the EMBL Data Library, September 1994
A|Reference number: S72664
A|Accession: S72664
A|Status: preliminary
A|Molecule type: mRNA
A|Residues: 1-140, 'C', 142-374 <KH2>
A|Cross-references: EMBL:X81695
C|Superfamily: Immunoglobulin C region; immunoglobulin homology

Query Match 36.7%; Score 1252.5; DB 2; Length 374;
Best Local Similarity 99.1%; Pred. No. 5, 9e-67;
Matches 229; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

395 EPKSCDKHTCPCPAPPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKF 454
143 EPKSCDKHTCPCPAPPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKF 202
455 NMVVDGVEVHNAKTKRREQYSTYRVSVLTITLHODWLNKGEYKCKVSNKALPAPIEKT 514
203 NMVVDGVEVHNAKTKRREQYSTYRVSVLTITLHODWLNKGEYKCKVSNKALPAPIEKT 262
515 ISKAGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTIP 574
263 ISKAGQPREPQVYTLPPSRDEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTIP 322
575 PVLDSGSEFLYSKLTVDKSRMQQGVFSCSVMHEALHNHYTQKSLSLSPG 625
323 PVLDSGSEFLYSKLTVDKSRMQQGVFSCSVMHEALHNHYTQKSLSLSPG 373

RESULT 7
PT0207
19 gamma chain C region - chimpanzee
C|Species: Pan troglodytes (chimpanzee)
C|Date: 23-Nov-1991 #sequence_revision 23-Nov-1991 #text_change 16-Jul-1999
C|Accession: PT0207
R|Ehrlich, P.H.; Moustafa, Z.A.; Oestberg, L.
Mol. Immunol. 28, 319-322, 1991
A|Title: Nucleotide sequence of chimpanzee Fc and hinge regions.
A|Reference number: PT0207; PMID:91287716; PMID:262315
A|Accession: PT0207
A|Molecule type: mRNA
A|Residues: 1-234 <EHR>
C|Superfamily: Immunoglobulin C region; immunoglobulin homology
C|Keywords: Immunoglobulin
F|48-117/Domain: Immunoglobulin homology <IM>

Query Match 35.7%; Score 1218; DB 2; Length 234;
Best Local Similarity 95.4%; Pred. No. 3, 6e-65;
Matches 227; Conservative 1; Mismatches 6; Indels 4; Gaps 1;

382 SNIKLPTWSTVEPKSCDKHTCPCPAPPELLGGPSVFLFPPKPKDTLMISRTPEVTCV 441
1 SNIKLPTWSTVEPKSCDKHTCPCPAPPELLGGPSVFLFPPKPKDTLMISRTPEVTCV 56
442 VVDVSHEDPEVKFMVVDGVEVHNAKTKRREQYSTYRVSVLTITLHODWLNKGEYKCK 501
57 VVDVSHEDPEVKFMVVDGVEVHNAKTKRREQYSTYRVSVLTITLHODWLNKGEYKCK 116

```

Query Match	34.2%	Score 1169	DB 2	Length 377
Query	502	VSNNKLPAPIEKTIISAKGQPREPOVYTLTPPSRDELTKNOVSLTCLVKGFPSPDIAVME	561	
Db	117	VSNNKLPAPIEKTIISAKGQPREPOVYTLTPPSRDELTKNOVSLTCLVKGFPSPDIAVME	176	
Qy	562	SNQGEENNYKTIIPVLDSDGSPFLYSKLTVDKSRMOQGNVFCSCVMHEALHNHYTKQS	619	
Db	177	SSQGEENNYKTIIPVLDSDGSPFLYSKLTVDKSRMOQGNVFCSCVMHEALHNHYTKQS	234	
RESULT 8				
A23511				
Ig gamma-3 chain C region (allotype G3m(b)) - human				
C:Species: Homo sapiens (man)				
C:Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 23-Jul-1999				
C:Accession: A23511				
R:Huck, S.; Fort, P.; Crawford, D.H.; Lefranc, M.P.; Lefranc, G.				
Nucleic Acids Res. 14, 1779-1789, 1986				
A:Title: Sequence of a human immunoglobulin gamma 3 heavy chain constant region gene: cc				
A:Reference number: A23511; MUID:86148507; PMID:3081877				
A:Accession: A23511				
A:Molecule type: DNA				
A:Residues: 1-377 <HUC>				
A:Cross-references: GB:X03604; GB:M12958; NID:G33070; PIDN:CAA27268.1; PID:G577056				
C:Genetics:				
A:Gene: GDB:IGHG3				
A:Cross-references: GDB:119339; OMIM:147120				
A:Map position: 14q32.33-14q32.33				
A:Introns: 98/3; 115/3; 130/3; 145/3; 160/3; 270/3				
C:Superfamily: immunoglobulin C region; immunoglobulin homology				
C:Keywords: immunoglobulin				
F:20-85/Domain: immunoglobulin homology <IMM>				
Query Match	34.3%	Score 1171	DB 2	Length 377
Best Local Similarity	92.6%	Pred. No. 4e-62		
Matches 214; Conservative 8; Mismatches 9; Indels 0; Gaps 0;				
Qy	395	EPKSCDKTKTPCPAPPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKF	454	
Db	146	EPKSCDTPPCRCRCPPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVQF	205	
Qy	455	NWYDGVGVNAKTKREBQVNSTYRVSVSLTVLHODMLNGKEYCKKVSNNKALPAPIEKT	514	
Db	206	KWYDGVGVNAKTKREBQVNSTYRVSVSLTVLHODMLNGKEYCKKVSNNKALPAPIEKT	265	
Qy	515	ISKAGQPREPOVYTLTPPSRDELTKNOVSLTCLVKGFPSPDIAVMESSNGPENNYKTTT	574	
Db	266	ISKTKGQPREPOVYTLTPPSREEMTKNOVSLTCLVKGFPSPDIAVMESSNGPENNYKTTT	325	
Qy	575	PVLDDSDGSPFLYSKLTVDKSRMOQGNVFCSCVMHEALHNHYTKQSLSPG 625		
Db	326	PVLDDSDGSPFLYSKLTVDKSRMOQGNVFCSCVMHEALHNHYTKQSLSPG 376		
RESULT 9				
A60764				
Ig gamma-3 chain C region, form LAT - human				
C:Species: Homo sapiens (man)				
C:Date: 14-May-1993 #sequence_revision 14-May-1993 #text_change 16-Jul-1999				
C:Accession: A60764				
R:Huck, S.; Lefranc, G.; Lefranc, M.P.				
Immunogenetics 30, 250-257, 1989				
A:Title: A human immunoglobulin IGHG3 allele (Gmb0, b1, c3, c5, u) with an IGHG4 convert				
A:Reference number: A60764; MUID:90007613; PMID:2571587				
A:Accession: A60764				
A:Status: Preliminary				
A:Molecule type: DNA				
A:Residues: 1-377 <HUC>				
C:Superfamily: immunoglobulin C region; immunoglobulin homology				
C:Keywords: immunoglobulin				
F:20-85/Domain: immunoglobulin homology <IMM>				

R.Milstein, C.; Frangione, B.
 Biochem. J. 121, 217-225, 1971
 A>Title: Disulfide bridges of the heavy chain of human immunoglobulin G2.
 A.Reference number: A90253; MUID:72033500; PMID:4940472
 A.Contents: annotation; myeloma protein S₂, disulfide bonds
 R.Frangione, B.; Milstein, C.; Pink, J.R.L.
 Nature 221, 145-148, 1969
 A>Title: Structural studies of immunoglobulin G.
 A.Reference number: A9157; MUID:69064124; PMID:5782707
 A.Contents: annotation; S₂, disulfide bonds
 C.Genetics:
 A.Gene: GDB:IGHG2
 A.Cross-references: GDB:119338; OMIM:147110
 A.Map position: 14q32.33-14q32.33
 C.Complex: An immunoglobulin heterotrimer subunit consists of two identical light (kappa) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a
 C.Superfamily: immunoglobulin C region; immunoglobulin homology
 C.Keywords: duplication; glycoprotein; heterotrimer; immunoglobulin
 F/20-85/Domain: immunoglobulin homology <IM1>
 F/133-203/Domain: immunoglobulin homology <IM2>
 F/239-306/Domain: immunoglobulin homology <IM3>
 F/14/Disulfide bonds: interchain (to light chain) #status experimental
 F/27-83,140-200,246-304/Disulfide bonds: #status experimental
 F/102,103,106,109/Disulfide bonds: interchain (to heavy chain) #status experimental
 F/176/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 33.7%; Score 1150; DB 1; Length 326;
 Best Local Similarity 88.9%; Pred. No. 5,8e-61;
 Matches 217; Conservative 9; Mismatches 10; Indels 8; Gaps 3;

QY 382 SNIKVLPTWSTVEPEKSCDKHTPCPAPELLGSPVLPFPKPKDTLMISRTPEVTCV 441
 DB 90 SNTKVDKT-----VERKCCVE---CPCCPAP-P-VAGPSVLPFPKPKDTLMISRTPEVTCV 141

QY 442 VVDVSHEDPEVKFNNYVDGEVYNAKTKREEQYNSTRVSVLTVLHDMINGKEYCK 501
 DB 142 VVDVSHEDPEVKFNNYVDGEVYNAKTKREEQYNSTRVSVLTVLHDMINGKEYCK 201

QY 502 VSNKLLPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNOVSLTCLVGFYPSDIAVEWE 561
 DB 202 VSNKGLPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNOVSLTCLVGFYPSDIAVEWE 261

QY 562 SNGQPENNYKTPPVLDSDGSFELYSKLTVDSKRNQOGNVSFCSVMHEALNHYTQKSL 621
 DB 262 SNGQPENNYKTPPVLDSDGSFELYSKLTVDSKRNQOGNVSFCSVMHEALNHYTQKSL 321

QY 622 LSPG 625
 DB 322 LSPG 325

RESULT 12
 A66254
 CD4 precursor - rabbit
 C.Species: Oryctolagus cuniculus (domestic rabbit)
 C.Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jan-2000
 C.Accession: A46254
 R.Hague, B.F.; Savadkoe, S.; Brown, T.J.; Lee, K.; Recker, D.P.; Kindt, T.J.
 Proc. Natl. Acad. Sci. U.S.A. 89, 7963-7967, 1992
 A>Title: CD4 and its role in infection of rabbit cell lines by human immunodeficiency virus
 A.Reference number: A46254; MUID:92390370; PMID:1518821
 A.Accession: A46254
 A.Status: preliminary
 A.Molecule type: mRNA
 A.Residues: 1-459 <HAG>
 A.Cross-references: GB:M92840; NID:9164871; PID:AAA1198.1; PID:9164872
 A.Note: sequence extracted from NCBI backbone (NCBIN:112732, NCBI:112733)
 C.Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
 F/322-372/Domain: immunoglobulin homology <IM1>

Query Match 33.7%; Score 1149; DB 2; Length 459;
 Best Local Similarity 56.9%; Pred. No. 1e-60;
 Matches 243; Conservative 66; Mismatches 98; Indels 20; Gaps 6;

QY 1 MNRGVPFRLHLLVLOALLPATQGNKVVLGKGPVETLCTASOKKSIOPHMKNSNOIK 60
 DB 1 MNRRIYFOCLLIVLPALLPATWKTIVYRKAGIVLPCOSSQKRNKSVNMKIANQV 60

QY 61 ILNGG-----SFLTKGPKLDRADSRSLWDQGNFPLIINKLKEDSYICEVEDQKE 116
 DB 61 ILNGGSSSSSPWLKGNPPLSNRVSKKMMMDQGSFPLVIDLRNDDSGTYICEVGDKKM 120

QY 117 EVQLLVFGLTANSDTHLLQGSLITLTLESPPSSSVCCSPKGNIGGKTLVSQLEL 176
 DB 121 EVELLVFRLTANPNRLRHGSLITLTLEPSPSVQMKSPENKIETGPTCSMPKRL 180

QY 177 QDSGWTCTV-LONQKVEPKIDIVLAFAKASSIVYKKEGQVFEFPLAFTVTKLGS 235
 DB 181 QDSGWSCHLSFQDNKLELDIKIIVLPKASATVYKKGQVFEFPLAFTVTKLGS 238

QY 236 GELMWQAERASSSKSWITFDLKNKESVYKATYDOPKLGKPLPLHLTPQALPOYAGSG 295
 DB 239 GELMWQVAGASAGSWSFSLDRKVSQKILPDLKIQMSGKPLSTLPLQALHRYAGSG 298

QY 296 NITLALAKTKGLHGVNLVVMRATQLOKNTLCEYWGFTSPKLMSTLENKEAKVSKRE 355
 DB 299 NITSLTLD--KGLIHQVSLVMLKTVQYKMKLTCEVLGPIDPKMKLSLKLIEDQEAQVS-TQ 355

QY 356 KPVWLVNPEAGMWOCILSDSGOVLLESNIKVLPTWSTVEPEKSCDKHTPCCPAPPELLG 415
 DB 356 KPVQVLDPEKAGWOCILSDSGOVLLESKADVLAT-----GLSHQPTLLAGALGG 405

QY 416 GPSVFLP 422
 DB 406 TAGLVLF 412

RESULT 13
 G4HU
 Ig gamma-4 chain C region - human
 C.Species: Homo sapiens (man)
 C.Date: 02-Apr-1982 #sequence_revision 02-Apr-1982 #text_change 16-Jul-1999
 C.Accession: A90933; A90249; A02150
 R.Ellison, J.; Buxbaum, J.; Hood, L.
 DNA 1, 11-18, 1981
 A>Title: Nucleotide sequence of a human immunoglobulin C-gamma4 gene.
 A.Reference number: A90933; MUID:83157104; PMID:6299662
 A.Accession: A90933
 A.Molecule type: DNA
 A.Residues: 1-327 <ELL>
 A.Note: the sequence was determined from the germ-line gene
 R.Pink, J.R.L.; Buttery, S.H.; De Vries, G.M.; Milstein, C.
 Biochem. J. 117, 33-47, 1970
 A>Title: Human immunoglobulin subclases. Partial amino acid sequence of the constant r
 A.Reference number: A90249; MUID:70207560; PMID:4192699
 A.Accession: A90249
 A.Molecule type: protein
 A.Residues: 1-30; 81-326 <PIN>
 C.Genetics:
 A.Gene: GDB:IGHG4
 A.Cross-references: GDB:119340; OMIM:147130
 A.Map position: 14q32.33-14q32.33
 A.Introns: 99/1; 111/1; 221/1
 C.Complex: An immunoglobulin heterotrimer subunit consists of two identical light (kappa) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a
 C.Superfamily: immunoglobulin C region; immunoglobulin homology
 C.Keywords: duplication; glycoprotein; heterotrimer; immunoglobulin
 F/20-85/Domain: immunoglobulin homology <IM1>
 F/133-203/Domain: immunoglobulin homology <IM2>
 F/240-307/Domain: immunoglobulin homology <IM3>
 F/14/Disulfide bonds: interchain (to light chain) #status experimental
 F/27-83,141-201,247-305/Disulfide bonds: #status predicted
 F/106,109/Disulfide bonds: interchain (to heavy chain) #status experimental
 F/177/Binding site: carbohydrate (Asn) (covalent) #status predicted

```

Query Match      33.4%; Score 1141; DB 1; Length 327;
Best Local Similarity 79.4%; Pred. No. 2e-60;
Matches 216; Conservative 16; Mismatches 22; Indels 18; Gaps 2;

QY 371 LLSSGOVLLESNIKVLPTWSTPVEPKSCDKTH-----TCPCPAPEL 413
D 56 VLGSSGLVSL--SSVTVVSSSLGRTYTCVNDHKRSNTKVDKRVESKXGPPCPAPEF 114
QY 414 LGSSVFLFPKPKDITLISRTPEVTCVVVSHEDPEVKFNMYVDGVEVNAATKPRE 473
D 115 LGSSVFLFPKPKDITLISRTPEVTCVVVSHEDPEVKFNMYVDGVEVNAATKPRE 174
QY 474 QYNSTYRVSVLTVLHQDMLNGEKYCKVSKALPAPIEKTISAKGQPREPQVYTLPPS 533
D 175 QFNSTYRVSVLTVLHQDMLNGEKYCKVSKALPAPIEKTISAKGQPREPQVYTLPPS 234
QY 534 RDELTKNOVSLTCLVKGFPSPDIAMWESNGOPENNYTTPPVLDSDSFFLYSKLTVDK 593
D 235 QEBMTKNOVSLTCLVKGFPSPDIAMWESNGOPENNYTTPPVLDSDSFFLYSKLTVDK 294
QY 594 SRMOQGVFSCSVNHEALHNHYTKSLSPG 625
D 295 SRMOQGVFSCSVNHEALHNHYTKSLSPG 326

RESULT 14
S30193
T-cell surface glycoprotein CD4 - dog
C:Species: Canis lupus familiaris (dog)
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 21-Jul-2000
C:Accession: S30193
R:Miller, K.F.; Conner, G.E.; Mintz, D.H.; Alejandro, R.
Biochim. Biophys. Acta 1172, 315-318, 1993
A:Title: Primary structure of the canine CD4 antigen.
A:Reference number: S30193; MUID:93192324; PMID:7916632
A:Accession: S30193
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-432 <ML>
A:Cross-references: EMBL:X68565; NID:G288652; PIDN:CA837664.1; PID:G4467377
C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
C:Keywords: glycoprotein
F:202-311/Domain: immunoglobulin homology <IMM>

Query Match      32.4%; Score 1105; DB 2; Length 432;
Best Local Similarity 57.4%; Pred. No. 3.8e-58;
Matches 229; Conservative 62; Mismatches 90; Indels 18; Gaps 6;

QY 12 LVLDLALIPATQGNKVVLGKGDVETLTCTASQKSIQFHWKNSNQIKITGNQSFVTK 71
D 1 LMLQLVMLPAVTPVREVLGKAGDAVELPCQTSQKNIHFNRDSSMWQIIGNQSFWTV 60
QY 72 GPKLNDRADRSRLMDQGNFPLIIKNLKIEDSDTYICEVDQKEVQLVPGTAA---- 127
D 61 GSSRLKHVESKKNLMDGSPFLVIKDELVADSGTYFCDT-DKQEVLLVFNLTAKWDS 119
QY 128 -----NSDTHLQSGSLTLTLESPPGSSPVQCRSPRKNIOGKTLTSLVSOLEIDSGT 182
D 120 GSSGSSNIRLLOQOQLTLTENPSSSPVQWKGPKGKSGONLSLSPHELQDGTW 179
QY 183 TCTVLQNGKVEFKIDIVLAFOKASSIVYKKEGQVEFSPLAFTVEKLTGSGELWQA 242
D 180 TCIISQSQKVEEFINIVLAFOKASSIVYKKEGQVEFSPLAFTVEKLTGSGELWQA 237
QY 243 ERASSSKSWITFDLKNKESVYKRVTDPKLQMGKKLPHLTLPOALPOYAGSGNLTALE 302
D 238 QGASSSLWISFTLENRLSKMEKAHAPKLLKMKESLPRFTLPOLSRVYASGLITLNL- 296
QY 303 AKTKLHQEVNLLVVRATQLOKNITCEVWGTSPKMLSLKLENKEAKVSRKRPVWLN 362
D 297 AK-ETVQEVNLLVVRANASSQNNLTCEVLTGPTSPBLTSLSLKQAAKVSQGLVWVVD 355
QY 363 PEAGMOCCLSDSGOVLLESNIKVLPTWSTPVEPKSCDK 401

```

```

D 356 PEGTWMOCCLSDXKXVLLASSLNW----SSPVVYKSWPK 390

RESULT 15
A27449
T-cell surface glycoprotein CD4 precursor - rat
N:Alternate names: W3/25 antigen
C:Species: Rattus norvegicus (Norway rat)
C>Date: 21-May-1988 #sequence_revision 21-May-1988 #text_change 21-Jan-2000
C:Accession: A27449; A35433
R:Clark, S.J.; Vetterli, W.A.; Barclay, A.N.; Gagnon, J.; Williams, A.F.
Proc. Natl. Acad. Sci. U.S.A. 84, 1649-1653, 1987
A:Title: Peptide and nucleotide sequences of rat CD4 (W3/25) antigen: evidence for derive
A:Reference number: A27449; MUID:87175535; PMID:3104900
A:Accession: A27449
A:Molecule type: mRNA
A:Residues: 1-457 <CL>
A:Cross-references: GB:M15768; NID:G203387; PIDN:AAA0901.1; PID:G203388
R:David, S.J.; Ward, H.A.; Puklavac, M.J.; Williams, A.C.; Williams, A.F.; Barclay, A.N.
J. Biol. Chem. 265, 10410-10418, 1990
A:Title: High level expression in Chinese hamster ovary cells of soluble forms of CD4 T
A:Reference number: A35433; MUID:90285164; PMID:2113054
A:Contents: annotation
C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
C:Keywords: glycoprotein; membrane protein; surface antigen
F:219-300/Domain: immunoglobulin homology <IMM>

Query Match      29.3%; Score 999; DB 2; Length 457;
Best Local Similarity 52.9%; Pred. No. 7.7e-52;
Matches 207; Conservative 61; Mismatches 115; Indels 8; Gaps 5;

```

```

QY 1 MNRGVPRRL--LLVQLALIPATQGNKVVLGKGDVETLTCTASQKSIQFHWKNSNQ 58
D 1 MCRGFSRHLPLLLDLSKLVVTVQKTYLVKEGSSALLPESTRRBASFMKSSDQ 60
QY 59 IKITGNQSFVTKGPKLNDRADRSRLMDQGNFPLIIKNLKIEDSDTYICEVDQKEV 118
D 61 KTLIGYKNKLLIGSLELYRSPSRKNAWERSGFPILINKLRMBDSQTYCELENKKEV 120
QY 119 QLVFGLTASDPHLQSGSLTLTLES-PPGSSPVQCRSPRKNIOGKTLTSLVSOLEQ 177
D 121 ELWVFRVTFPGRLLQSGSLTLTLDSPVSDPPECKKSSNIVDSKAFTHSLRIQ 180
QY 178 DSGTWTCTVLQNGKVEFKIDIVLAFOKASSIVYKKEGQVEFSPLAFTVEKLTGSGE 237
D 181 DSGTWTCTVLQNGKVSFDMKLSVLGFASTISITAYKSEGSABSFPLNAGEBSL--QGE 238
QY 238 LRMKAERAPSSQWITFSLNKNQKVSQKSTSNKPFQLEFLLPTLIQIPVSLGPAAGSGL 298
D 298 TLLAEATGTLHDEVNLLVVRATQLOKN-LTCEVWGTSPKMLSLKLENKEAKVSRK 356
QY 299 TLLTD--RGLYQEVNLLVVRATQLOKNTLTCEVWGTSPKMLSLKLENKEAKVSRK 356
D 357 PAVVNLPEAGMOCCLSDSGOVLLESNIKVL 387
D 357 VIQVQAPPEAGMOCCLSDSGEVEKMDSKIVL 387

RESULT 16
RMNST4
T-cell surface glycoprotein CD4 precursor - mouse
N:Alternate names: T-cell differentiation antigen L3T4; T-cell surface antigen T4/Leu 3
C:Species: Mus musculus (house mouse)
C>Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 16-Jul-1999
C:Accession: A02110; A26038; A39953; A39955; I54564; I59018; A47642
R:Tourville, B.; Gorman, S.D.; Field, E.H.; Hunkapiller, T.; Parnes, J.R.
Science 234, 610-614, 1986
A:Title: Isolation and sequence of L3T4 complementary DNA clones: expression in T cells
A:Reference number: A02110; MUID:87018845; PMID:3094146
A:Accession: A02110

```

A:Molecule type: mRNA
 A:Residues: 1-457 <TOU>
 A:Cross-references: GB:M13816; NID:G192070; PIDN:AAA37267.1; PID:G309112
 R:Litman, D.R.; Gettner, S.N.
 Nature 325, 453-455, 1987
 A:Title: Unusual intron in the immunoglobulin domain of the newly isolated murine CD4 (IL3) cDNA
 A:Reference number: A26038; MUID:87115821; PMID:3027575
 A:Accession: A26038
 A:Molecule type: mRNA
 A:Residues: 1-457 <LIT>
 A:Cross-references: GB:X04836; NID:G50353; PIDN:CAA28539.1; PID:G50354
 R:Gorman, S.D.; Tourville, B.; Parne, J.R.
 Proc. Natl. Acad. Sci. U.S.A. 84, 7644-7648, 1987
 A:Title: Structure of the mouse gene encoding CD4 and an unusual transcript in brain.
 A:Reference number: A39893; MUID:88041159; PMID:2823269
 A:Accession: A39893
 A:Molecule type: DNA
 A:Residues: 1-25, 'E', 27-457 <GOR>
 A:Cross-references: GB:M17080; GB:J03003; NID:G192515; PIDN:AAA37402.1; PID:G387124
 R:Maddon, P.J.; Mollineux, S.M.; Maddon, D.E.; Zimmerman, K.A.; Godfrey, M.; Alt, F.W.; Proc. Natl. Acad. Sci. U.S.A. 84, 9155-9159, 1987
 A:Title: Structure and expression of the human and mouse T4 genes.
 A:Reference number: A39955; MUID:88097446; PMID:3501122
 A:Accession: A39955
 A:Status: nucleic acid sequence not shown; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 25-457 <MAD>
 A>Note: the cited Genbank accession number, J03564, is not in release 101.0
 R:Parne, J.R.; Hunkapiller, T.
 Immunol. Rev. 100, 109-127, 1987
 A:Title: L3T4 and the immunoglobulin gene superfamily: New relationships between the immunoglobulin and T-cell receptor genes.
 A:Reference number: 154564; MUID:88152875; PMID:3326818
 A:Accession: 154564
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-457 <RES>
 A:Cross-references: GB:M36850; NID:G198670; PIDN:AAA39401.1; PID:G198671
 A:Accession: 169018
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 208-318 <RE2>
 A:Cross-references: GB:M36851; NID:G198672; PIDN:AAA39402.1; PID:G554183
 R:Classon, B.J.; Teegarden, J.; Kitzbaum, L.; Maddox, J.; Mackay, C.R.; Brandon, M.; M Immunogenetics 23, 129-132, 1986
 A:Title: The L3T4 antigen in mouse and the sheep equivalent are immunoglobulin-like.
 A:Reference number: A47642; MUID:86166694; PMID:3082751
 A:Accession: A47642
 A:Molecule type: protein
 A:Residues: 27-43 <CLA>
 A:Comment: This protein is expressed on most thymocytes, on a subset of mature T-cells
 C:Genetics:
 A:Introns: 18/1; 74/1; 128/1; 207/1; 319/1; 386/1; 425/3; 448/2
 C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
 C:Keywords: alternative initiators; duplication; glycoprotein; T-cell; transmembrane protein
 F:1-26/Domain: signal sequence #status predicted <Sig>
 F:1-27/Domain: signal sequence #status predicted <Sig>
 F:27-457/Domain: T-cell surface glycoprotein CD4 #status experimental <MAT>
 F:35-114/Domain: immunoglobulin homology <IM1>
 F:119-190/Domain: immunoglobulin homology #status atypical <IM2>
 F:220-301/Domain: immunoglobulin homology <IM3>
 F:241-457/Domain: CD4, brain-specific short form #status predicted <BRA>
 F:321-372/Domain: immunoglobulin homology <IM4>
 F:395-419/Domain: transmembrane #status predicted <TM>
 F:420-457/Domain: intracellular #status predicted <INT>
 F:42-112, 159-188, 328-370/Disulfide bonds: #status predicted
 F:187, 298, 323, 392/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 29.1%; Score 993; DB 1; Length 457;
 Best Local Similarity 53.8%; Pred. No. 1.7e-51;
 Matches 211; Conservative 62; Mismatches 109; Indels 10; Gaps 7;

QY 1 MNRGVFRH-LILVQLALLPAPAGNKVGLGKGGDTVELTCTASQKSIQPHMNKSNQI 59
 DB 1 MCRATSLRLLTLTLQLSGLAVTGGKTLVLEKGGESALPESSQKTLTFWKFSDQR 60

QY 60 KILGNQG-SPLTKG--PSKLNDRADSRSLMDQGNFPIIKNLIKIEDSDTYICEVEDQKE 116
 DB 61 KILGQHGKGVLRGSSPSQF-DREFDSKGAWEKSFPIIKNLIKIEDSDTYICELENKE 119
 QY 117 EVQLVFLGFLTPSDTHLQGGSLTLTLLES-PPGSSPSVQCRSPRKNIOGQKTLSSVQLE 175
 DB 120 EVELWVFKVTPSPGTSILQGGSLTLTLSDNSKVSPLRECKHKKGVSSGKVSLSMNLK 179
 QY 176 LODSGTWCTVLQNGKVEPKIDIVLAFQKASSIVYKKEGQVFPPLAFTVEKLTGS 235
 DB 180 VQDSFPMKCTVTLDDKQKWFGLTSLVLFQSTALYKXSGESAFEPPLFAEB--NGW 237
 QY 236 GELTMQARASSSKSKWTFPDDKKNKVSRYKATYQDPLQMGKLPILTLTPQALPOLYASG 295
 DB 238 GELTMKAEKDSFPQWISFSTINKSVQSKTKDKLQKETPLTLTKIPOVSLQPAASG 297
 QY 356 KPVVTLNPDAGMWCCLLSDSGOVLESNIKVL 387
 DB 356 KVVQVVAPEETGLMQLTSEGDVKVMDSRIOVL 387

RESULT 17
 G3MSM
 Ig gamma-3 chain C region, membrane-bound form - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 13-Aug-1986 #sequence_revision 31-Mar-1991 #text_change 16-Jul-1999
 C:Accession: A02156; A02155
 R:Wels, J.A.; Word, C.J.; Rimm, D.; Der-Balan, G.P.; Martinez, H.M.; Tucker, P.W.; Blatt EMBO J. 3, 2041-2046, 1984
 A:Title: Structure analysis of the murine IgG3 constant region gene.
 A:Reference number: A02156; MUID:85027161; PMID:6092053
 A:Accession: A02156
 A:Molecule type: DNA
 A:Residues: 1-398 <WEL>
 A:Cross-references: GB:J00451; NID:G194392; PIDN:AB59655.1; PID:G194433
 A>Note: the sequence was determined from the germline gene
 R:Komaromy, M.; Clayton, L.; Rogers, J.; Robertson, S.; Kettman, J.; Wall, R. Nucleic Acids Res. 11, 6775-6785, 1983
 A:Title: The structure of the mouse immunoglobulin in gamma-3 membrane gene segment.
 A:Reference number: A02155; MUID:84041483; PMID:6314258
 A:Accession: A02155
 A:Molecule type: DNA
 A:Residues: 328-332, 'G', 334-341, 'Q', 343-387, 'F', 389-398 <KOM>
 A:Cross-references: GB:X00688
 A>Note: the sequence was determined from the germline gene
 C:Genetics:
 A:Introns: 97/1; 113/1; 223/1; 328/1; 371/3
 C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kap) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into a C:Superfamily: immunoglobulin C region; immunoglobulin homology
 C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobulin
 F:119-83/Domain: immunoglobulin homology <IM1>
 F:197-112/Region: hinge
 F:136-205/Domain: immunoglobulin homology <IM2>
 F:242-309/Domain: immunoglobulin homology <IM3>
 F:346-362/Domain: transmembrane #status predicted <TM>
 F:363-398/Domain: intracellular #status predicted <INT>
 F:179, 322/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 27.8%; Score 947.5; DB 1; Length 398;
 Best Local Similarity 60.1%; Pred. No. 7.2e-49;
 Matches 172; Conservative 44; Mismatches 49; Indels 21; Gaps 2;

QY 382 SNIKVLPWSPVPEKSCDKTH-----TCPP---CPAELLGGSPVF 420
 DB 64 SLLVVPSPSTWSPQVINCVAHPASKTELIRIEPRIPKSTPPSSSCPPIIGLGPSPVF 123
 QY 421 LFPPKXDTLMSIRPEVTCVVVDVSHEDPEVKFNVYDGVNNAKTKPREEQYNSTYR 480

```

Db      124 IFPPKXALMISLTPKKTCCVVDVSEDDPDVHVSFMDNKEVHTAMQPREADYNSTFR 183
      481 VSVSLTVLHODMLNGKEKCYKSNKALPAPIEKTISKAKGQPREPVLTLPSSDELTKN 540
      184 VVSLALPIQHQMKGKCKCYKSNKALPAPIEKTISKAKGQAPQPVTTTTPPEQNSKK 243
Qy      541 QVSLTCLVKGFFVPSDIAVWESNQGPNNTKTPPEVLDSDGSPFLYSKLTVDKSRWQGN 600
      244 KVSILTCLVNTFFSEALISYEMERNGELFEDYKNTPTILSDGTFFLYSKLTVDTSWLGGE 303
Qy      601 VFSGSVHAEALHNHTQKSLSLSPQLDFTCAEAQDGLGLMTT 646
      304 IFTGSVHAEALHNHTQKSLSLSPQLDFTCAEAQDGLGLMTT 349

```

RESULT 18

```

GIMSK
Ig gamma-1 chain C region, membrane-bound form - mouse
C/Species: Mus musculus (house mouse)
C/Date: 17-Dec-1982 #sequence revision 31-Mar-1991 #text_change 16-Jul-1999
C/Accession: B02159; MUID:80045036; PMID:115593
R/Honjo, T.; Ohta, M.; Yamawaki-Kataoka, Y.; Kataoka, T.; Kawakami, T.; Takahashi, N.;
Cell 18, 559-568, 1979
A/Title: Cloning and complete nucleotide sequence of mouse immunoglobulin gamma1 chain g
A/Reference number: A02159; MUID:80045036; PMID:115593
A/Accession: B02159
A/Molecule type: DNA
A/Residues: 1-393 <HON>
A/Cross-references: GB:J00453
A/Note: The sequence was determined from the germ-line gene
R/Tyler, B.M.; Cowman, A.F.; Geronakakis, S.D.; Adams, J.M.; Bernard, O.
Proc. Natl. Acad. Sci. U.S.A. 79, 2008-2012, 1982
A/Title: mRNA for surface immunoglobulin gamma chains encodes a highly conserved transme
A/Reference number: A02160; MUID:82197626; PMID:6804950
A/Accession: A02160
A/Molecule type: mRNA
A/Residues: 323-393 <TYL>
R/Rogers, J.; Choi, E.; Souza, L.; Carter, C.; Word, C.; Kuehl, M.; Eisenberg, D.; Wall,
Cell 26, 19-27, 1981
A/Title: Gene segments encoding transmembrane carboxyl termini of immunoglobulin gamma c
A/Reference number: A02158; MUID:82115295; PMID:6799207
A/Accession: B02158
A/Molecule type: DNA
A/Residues: 323-366 <ROG>
A/Note: this sequence is the translation of the first exon of the M segment
C/Comment: Cell lines producing IgG contain two mRNA species for Ig gamma chains. The me
nucleotide membrane-bound chains in that it contains an alternative 3' end, encoded in separa
C/Genetics:
A/Introns: 1/1; 98/1; 111/1; 218/1; 323/1; 366/3
C/Complex: An immunoglobulin heterotrimer subunit consists of two identical light (kap
chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la
C/Superfamily: immunoglobulin C region; immunoglobulin homology
C/Keywords: alternative splicing; duplication; glycoprotein; heterotrimer; immunoglob
F/131-200/Domain: immunoglobulin homology <IMM>
F/340-357/Domain: transmembrane #status predicted <TMM>
F/358-393/Domain: intracellular #status predicted <INT>
F/174,278/Binding site: carbohydrate (Asn) (covalent) #status predicted

```

```

Query Match      27.5% Score 940.5; DB 1; Length 393;
Best Local Similarity 63.2%; Pred. No. 1.8e-48;
Matches 165; Conservative 46; Mismatches 45; Indels 5; Gaps 2;

```

```

Qy      388 PTWSTPYEPKSCDKTHNCP--CPAPELLGSPVFLFPKPKDTLMISRPEVTCVVDV 445
      87 PASSTKDKKTLVPRDCKGKPCICTVPEV--SSVFIFPKPKDVLITLTPTKVCVVDI 143
Qy      446 SHEDPEVKFNMYVDGVEVHNAKTKPREQYNSTYRVVSLTVLHODMLNGKEKCYKSNK 505
      144 SKDDPEVQFSWFDVDEVHTAQTPREQYNSTFRVSSELPIMQDMLNGKEKCYKSNK 203
Qy      506 ALPAPIEKTISKAKGQPREPVLTLPSSDELTKNQVSLTCLVKGFFVPSDIAVWESNQG 565

```

```

Db      204 AFPAPIEKTISKTKGRKPAQVYTTTPPEKQMAKDKVSLTCMITDPEPDIIVEMQNGQ 263
Qy      566 PENNYKTTPTVLDSDGSPFLYSKLTVDKSRWQGNVPSCSVMHAEALHNHTQKSLSPG 625
      264 PAENYKTKQPIYNTNGSYFFYSKLTNQGKSNMEAGNTTCSVLHGLNHNHTKSLSPG 323
Qy      626 LQDFTCAEAQDGLGLMTT 646
      324 LQDFTCAEAQDGLGLMTT 344

```

RESULT 19

```

GHRB
Ig gamma chain C region - rabbit
C/Species: Oryctolagus cuniculus (domestic rabbit)
C/Date: 24-Apr-1984 #sequence revision 15-Nov-1984 #text_change 16-Jul-1999
C/Accession: A91749; A90290; A93928; A90245; A94416; A02161
R/Bernstein, K.E.; Alexander, C.B.; Mage, R.G.
Immunogenetics 18, 387-397, 1983
A/Title: Nucleotide sequence of a rabbit IgG heavy chain from the recombinant F-1 haplocl
A/Reference number: A91749; MUID:84030930; PMID:6313520
A/Accession: A91749
A/Molecule type: mRNA
A/Residues: 1-323 <BER>
A/Note: This sequence has the d12 allotypic marker, 104-Thr, and the e14 marker, 185-Thr
R/Pratt, D.M.; Mole, L.E.
Biochem. J. 151, 337-349, 1975
A/Title: Sequence studies on the constant region of the Fd sections of rabbit immunoglob
A/Reference number: A90290; MUID:76135465; PMID:1243651
A/Accession: A90290
A/Molecule type: protein
A/Residues: 1-47, 'E', '49-71', 'PV', '72-128 <PRA>
R/Martens, C.L.; Moore, K.W.; Steinmetz, M.; Hood, L.; Knight, K.L.
Proc. Natl. Acad. Sci. U.S.A. 79, 6018-6022, 1982
A/Title: Heavy chain genes of rabbit IgG: isolation of a cDNA encoding gamma heavy chain
A/Reference number: A93928; MUID:8329917; PMID:6193512
A/Accession: A93928
A/Molecule type: mRNA
A/Residues: 88-103, 'M', '105-143', 'E', '145-184', 'A', '186', 'E', '188-266 <MAR>
A/Cross-references: GB:M16426; NID:9165111; PIDN:AA31289.1; PID:9165112
A/Note: this sequence has the d11 allotypic marker, 104-Met, and the e15 allotypic marker
R/Fruehrer, R.G.; Jackson, S.A.; Mole, L.E.; Porter, R.R.
Biochem. J. 116, 249-259, 1970
A/Title: Sequence studies of the Fd section of the heavy chain of rabbit immunoglobulin (
A/Reference number: A90245; MUID:70110015; PMID:5461106
A/Accession: A90245
A/Molecule type: protein
A/Residues: 132-143, 'E', '145-161 <FRU>
R/Hill, R.U.; Lebovitz, H.E.; Fellows Jr., R.E.; Delaney, R.
in Gamma Globulins, Nobel Symp. 3, Killander, J., ed., pp.109-127, Almqvist and Wiksell,
A/Reference number: A94416
A/Accession: A94416
A/Molecule type: protein
A/Residues: 129-131, 'S', '135-172', 'D', '174-184', 'A', '186', 'E', '188-200', 'D', '202-217', 'E', '219-232', 'Q', '
C/Complex: An immunoglobulin heterotrimer subunit consists of two identical light (kap)
chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la
C/Superfamily: immunoglobulin C region; immunoglobulin homology
C/Keywords: duplication; glycoprotein; heterotrimer; immunoglobulin
F/20-82/Domain: immunoglobulin homology <IMM>
F/130-199/Domain: immunoglobulin homology <IM2>
F/236-303/Domain: immunoglobulin homology <IM3>
F/173/Binding site: carbohydrate (Asn) (covalent) #status predicted

```

```

Query Match      27.0% Score 922; DB 1; Length 323;
Best Local Similarity 69.3%; Pred. No. 1.8e-47;
Matches 169; Conservative 30; Mismatches 37; Indels 8; Gaps 2;

```

```

Qy      382 SNIKVPTWSTPYEPKSCDKTHNCP--CPAPELLGSPVFLFPKPKDTLMISRPEVTCV 441
      87 TTKTKVDKT-----VAPSTCKR-----PTCPPPELLGSPVFLFPKPKDTLMISRPEVTCV 138
Qy      442 VVDVSHEDPEVKFNMYVDGVEVHNAKTKPREQYNSTYRVVSLTVLHODMLNGKEKCK 501

```


RESULT 27

S22080

Ig heavy chain precursor (B/MT.4A.17.H5.A5) - bovine
N/Alternate names: Ig gamma-1 chain C region (clone 8.10)
C/Species: Bos primigenius taurus (cattle)
C/Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text_change 23-Jul-1999
C/Accession: S22080; S06610; A31303

R/Sanders, P.G.
submitted to the EMBL Data Library, November 1991
A/Reference number: S22080

A/Accession: S22080
A/Status: preliminary

A/Molecule type: mRNA
A/Residues: 1-470 <SN>

A/Cross-references: EMBL:X62916; NID:9439; PID:CAA44699.1; PID:9440
R/Symons, D.B.A.; Clarkson, C.A.; Beale, D.
Mol. Immunol. 26, 841-850, 1989

A/Title: Structure of bovine immunoglobulin constant region heavy chain gamma 1 and gamma 2
A/Reference number: S06610; MUID:90097956; PMID:2513487
A/Accession: S06610

A/Molecule type: DNA

A/Residues: 142-470 <SYM>
A/Cross-references: EMBL:X16701

A/Note: the sequence was determined from the germline gene
C/Genetics:

A/Gene: Ig CH gamma-1
A/Introns: 98/1; 111/1; 221/1

C/Superfamily: immunoglobulin C region; immunoglobulin homology
C/Keywords: glycoprotein; heterotetramer; immunoglobulin; membrane protein

F/161-225/Domain: immunoglobulin homology <IMM>
F/318/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 25.9%; Score 884.5; DB 2; Length 470;

Best Local Similarity 35.6%; Pred. No. 4.8e-45;
Matches 227; Conservative 58; Mismatches 164; Indels 189; Gaps 16;

```

Oy 6 PFRHLVLVQ-----LALLPAATQGNKVVLAGKGGDTVLTCTAS--QKSIQFMKSNQ 58
Db 3 PLMTLLFLSLAPIGVLSGVQVRESGPSLV--KPSQTLSTCTVVGFSLSALTWVROAP 60
Oy 59 IKILGNOSFLTKGPSKLNDRADSRSLMDGN--FPLIIKNLKIEDSDTYIC----- 109
Db 61 GKALEWVGITSGGTYTYPALKSRLSTIKENSKSVLSVSTPBDTATYTCARSTYG 120
Oy 110 EVEDQKEVQLVLGTLTANSTHLLQGSLLTLESPPGSSPSVQCSPRKNIQGGKTL 169
Db 121 EVGG-----GAIADA-----WGGGLLVTVSSASTPAKYPLS---SCCGKSS 161
Oy 170 SVSOLLEQDSGTWCTVVLQNGKVEFKIDIVLAFQKASSIVYKKEGQVEFSPPLAFTV 229
Db 162 STVTL-----GCLVSYNPEPTV 180
Oy 230 EKLTSGLMLWQABRASSKSMITFDLNKKEVSVKRVYTDPKLQWKKLPLHTLPLPALP 289
Db 181 -----TMSGALK----- 188
Oy 290 QYASSGNLTALAEKTKGLHBEVNLVWRARATQOLKNTLCEVWGPSPKMLSLKLEKKA 349
Db 189 ----SGVTFPAVLQSSGLYSLSLNMVTPGSGTQTFCNVAHPAS-----ST 232
Oy 350 KVSREKPVVWLNPEAGMWQCLLSDSGVLLSNIKVLPTWSTVPEPKSCDKHTCPPCP 409
Db 233 KVDAAVDP-----CKSPCC-----C-CP 251
Oy 410 APELLGGSVPLFPKPKDQTLMIKSTPEVTCVVVDVSHEDPEVKFNWYVDGVEVNAKTK 469
Db 252 PPELPGGSPVIFPPKPKDTLTISGTPEVTCVVVDVGHDDPEVFNWYVDGVEVNAATTK 311
Oy 470 PREQVNSTYVSVLVTLHODWMLNGKSKYKCNKALPAIEKTIISAKQPREPQVYT 529
Db 312 PREQVNSTYVSVVSLARIQHDWVGKFKCKVHNEGHPAIVITISRTKGPAREPQVYV 371

```

RESULT 28

147161

Ig gamma 3 chain constant region - pig (fragment)
C/Species: Sus scrofa domestica (domestic pig)
C/Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 21-Jan-2000
C/Accession: 147161

R/Kaczkovics, I.; Sun, J.; Butler, J.B.
J. Immunol. 153, 3565-3573, 1994

A/Title: Five putative subclasses of swine IgG identified from the cDNA sequences of a
A/Reference number: 147158; MUID:95015845; PMID:7930579

A/Accession: 147161
A/Status: preliminary; translated from GB/EMBL/DDBJ

A/Molecule type: mRNA
A/Residues: 1-328 <KAC>

A/Cross-references: EMBL:U03781; NID:9433127; PID:AAA52219.1; PID:9433128
C/Genetics:

A/Gene: IgG3
C/Superfamily: immunoglobulin C region; immunoglobulin homology

F/133-202/Domain: immunoglobulin homology <IMM>

Query Match 25.9%; Score 883.5; DB 2; Length 328;
Best Local Similarity 62.0%; Pred. No. 3.4e-45;

Matches 170; Conservative 32; Mismatches 51; Indels 21; Gaps 4;

```

Oy 371 LLSDSGVLLSNIKVLPTWSTVPEPKSCDKHT-----TCPECPAPEL 413
Db 56 VLQPSGLYSLSMVTV--PASLSKSKYTCNVNHPATTYKDKRVGTRKPPCPICPCE- 113
Oy 414 LGGSPVLPFPKPKDQTLMIKSTPEVTCVVVDVSHEDPEVKFNWYVDGVEVNAKTKPRE 473
Db 114 VAGPSVIFPPKPKDTLMISGTPEVTCVVVDVSHEDPEVKFNWYVDGVEVNAATRKRE 173
Oy 474 QYNSTYVSVLVTLHODWMLNGKSKYKCNKALPAIEKTIISAKQPREPQVYTLPSS 533
Db 174 QPNSTYVSVLVTLHODWMLNGKSKYKCNKALPAIEKTIISAKQPREPQVYTLPSP 233
Oy 534 RDELTKQVSLTCLVKGFPYSDIAVEMESNGQ--PENNYKTTTPVLDSDGSPFYSLTLV 591
Db 234 AEELSRKVTVTCLVIGFYPPDIHVEKNSNGQPEPEGNRYRTTPQDDVDGTFLLYSKLAV 293
Oy 592 DKSRWQGNVFCSCVMHEALHNHYTKSLSPG 625
Db 294 DKARMDHGETFECAVMHEALHNHYTKSLSTQG 327

```

RESULT 29

S31459

Ig gamma-1 chain - sheep (fragment)
C/Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C/Accession: S31459

R/Patri, S.; Nau, F.
submitted to the EMBL Data Library, December 1992
A/Reference number: S31459

A/Accession: S31459
A/Status: preliminary

A/Molecule type: mRNA
A/Residues: 1-472 <PAT>

A/Cross-references: EMBL:X69797
C/Superfamily: immunoglobulin C region; immunoglobulin homology

C/Keywords: immunoglobulin

F/277-346/Domain: immunoglobulin homology <IMM>

A:Accession: P50018
A:Molecule type: DNA
A:Residues: 1-333 <BRU>
R:Brueggemann, M.; Free, J.; Diamond, A.; Howard, J.; Cobbold, S.; Waldmann, H.
Proc. Natl. Acad. Sci. U.S.A. 83, 6075-6079, 1986
A>Title: Immunoglobulin heavy chain locus of the rat: striking homology to mouse antibody
A:Reference number: A55941; MUID:86287397; PMID:3016742
A:Accession: B25941
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 227-333 <BR2>
C:Genetics:
A:Introns: 96/1; 117/1; 227/1
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: immunoglobulin
F:20-82/Domain: immunoglobulin homology <IMM>

Query Match 24.7%; Score 844; DB 2; Length 333;
Best Local Similarity 61.7%; Pred. No. 7.6e-43;
Matches 153; Conservative 34; Mismatches 51; Indels 10; Gaps 1;

QY PTWSTPVEPK-----SCDKTHTCPCPAPBLGGPSVFLPPPKXTLMTISRPE 437
DB PASSTKVKKVBRNNGIGHCPTCTCHKCPVPLGGPSVFLPPPKDILISQAK 144
QY 438 VTCVVVDVSHEDPEVKFNNYV DGVENNAKT KPREEOYNSTYRVSVLT VLVHODM LNKGE 497
DB 145 VTCVVVDVSEEDPVQF SFWFN NVEVHTAQTQPREQYNSTFRVYSA LPIQHODM SGE 204
QY 498 YKCKVSKALPAPIEKITISKAKGPREQVYTLPPSRDELTKNOVSLTCLVKGYPSPDIA 557
DB 205 FKCKNNALPSPLEIKTISKRGGLVKKPQVYVVGPTQLTEQVLSLCLTSGFLPNDIG 264
QY 558 VEWESNGQPENNYKTPPVLDSDGSFPLYSKLTVDKSRQGNVSCGVMHEALHNHYTQ 617
DB 265 VEWESNGIEKQYKTEPRVMDSDGSFMYSKLVNRSKMSDRAPVCSVHIEGLHNHVE 324
QY 618 KSLSLSPG 625
DB 325 KSLSRPFG 332

RESULT 33
PC4436
monoclonal antibody 13-1 heavy chain - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 04-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 21-Jan-2000
C:Accession: PC4436
R:Akashi, S.; Kato, K.; Torizawa, T.; Dolmae, N.; Yamaguchi, H.; Kamachi, M.; Harada, A.
Biochem. Biophys. Res. Commun. 240, 566-572, 1997
A>Title: Structural characterization of mouse monoclonal antibody 13-1 against a porphyrin
A:Reference number: J5810; MUID:98063277; PMID:9398605
A:Accession: PC4436
A:Molecule type: protein
A:Residues: 1-444 <AKA>
C:Comment: This catalytic antibody has peroxidase oxidase activity. It is directed against
F:251-330/Domain: immunoglobulin C region; immunoglobulin homology <IMM>
F:122/Diulfide bonds: interchain (to 98) #status predicted
F:199/Diulfide bonds: interchain (to 109) #status predicted

Query Match 24.6%; Score 839.5; DB 2; Length 444;
Best Local Similarity 60.8%; Pred. No. 2e-42;
Matches 146; Conservative 44; Mismatches 45; Indels 5; Gaps 2;

QY 388 PTWSTPVEPKSCDKTHTCPP--CPAPBLGGPSVFLPPPKXTLMTISRPEVTCVVVDV 445
DB 207 PASSTKVKKVBRNNGIGHCPTCTCHKCPVPLGGPSVFLPPPKDILISQAK 144
QY 446 SHEDPEVKFNNYV DGVENNAKT KPREEOYNSTYRVSVLT VLVHODM LNKGEYKCKVSK 505
DB 264 SKDPEVQFSWFDVDEVHTAQTQPREQYNSTFRVYSA LPIQHODM LNKGEYKCKVSK 323

QY 506 ALPAPIEKITISKAKGPREQVYTLPPSRDELTKNOVSLTCLVKGYPSPDIWESNGQ 565
DB 324 AFPAPIEKITISKGRKAPQVYTLPPPKQMKADKSLTCTMTIDFPEDITVEMQNNQ 383
QY 566 PENNYKTPPVLDSDGSFPLYSKLTVDKSRQGNVSCGVMHEALHNHYTQKSLSPG 625
DB 384 PAENYKTPQIMDTDSGFYVSKLVNRSKMSDRAPVCSVHIEGLHNHTEKSLSPG 443

RESULT 34
S37483
Ig gamma-2a chain - mouse
C:Species: Mus musculus (house mouse)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 23-Jul-1999
C:Accession: S37483
R:Duncanson, F.F.D.
submitted to the EMBL Data Library, February 1993
A:Reference number: S37483
A:Accession: S37483
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-469 <DUC>
A:Cross-references: EMBL:X70423; NID:g406252; PID:CA49868.1; PID:g406253
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: immunoglobulin
F:276-345/Domain: immunoglobulin homology <IMM>

Query Match 24.5%; Score 836.5; DB 2; Length 469;
Best Local Similarity 35.4%; Pred. No. 3.3e-42;
Matches 216; Conservative 66; Mismatches 142; Indels 187; Gaps 21;

QY 30 LGKKGDVLELTCTASQKSIQFPMKNSQIKLGNQ-----SFLTKGSKLNDRAISR 83
DB 30 LVKPGASVKSICRASGVTFTDY--INWYKQKPGQGLKMGITYPASGNTKYNENK 86
QY 84 RSL---WDGNNFLLIKNLKIEDSDTYICEVDQKEVQLVFGLTANSSTHLIQ--GQS 138
DB 87 ATLIVDTSSSTAYWQLSLTSEDTAVVFC-----ARANGATAT---LLDYWGQG 132
QY 139 LTLTLSPSPSPVOGRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLQNKVVERKID 198
DB 133 TLTLTSSAKTTAPSVPLAP---VCGDTT----- 158
QY 199 IVVLAQKASSIYKKEGEQVEFSPLAFTVEKLTGSGELMQAERASSSKSWITPDLKN 258
DB 159 -----GSSVTL---GCLVKGYPF---EYVT---LTWNSGSLSSG----- 188
QY 259 KEYSVKRVTPDKLQNGKCLPLHLTLPLQALPQ--YAGSGNLTALAEKTKGHOENVLV 316
DB 189 -----VH--TEPAVLAQSDLYTLSSSVT-----V 209
QY 317 MRATQLOKNTLCYWGPTSPKMLSLKLEKKAIVSRKRPVWVLYNPEAGMOCCLSDSG 376
DB 210 TSSTPSQSITCVNAPHAS-----STKVDKTEP-----RG 240
QY 377 QVLESNIKLTPTWSTPVEPKSCDKTHTCPP--CPAPBLGGPSVFLPPPKXTLMTISR 434
DB 241 -----PT-----IKP-----CPKCPAPNLLGGPSVFLPPKIDVLMISL 277
QY 435 TPBYTCVVVDVSHEDPEVKFNNYV DGVENNAKT KPREEOYNSTYRVSVLT VLVHODM L 494
DB 278 SPVITCVVDVSDSDPDVQISWFN NVEVHTAQTQPREQYNSTLRVYSA LPIQHODM S 337
QY 495 GKXKCVSKALPAPIEKITISKAKGPREQVYTLPPSRDELTKNOVSLTCLVKGYPSP 554
DB 338 GKXKCVSKALPAPIEKITISKAKGPREQVYTLPPSRDELTKNOVSLTCLVKGYPSP 397
QY 555 DIWESNGQPENNYKTPPVLDSDGSFPLYSKLTVDKSRQGNVSCGVMHEALHNH 614
DB 398 DIWESNGQPENNYKTPPVLDSDGSFPLYSKLTVDKSRQGNVSCGVMHEALHNH 457
QY 615 YTKSLSLSPG 625

Db 458 HTKSFSPRTFG 468

RESULT 35

PS0017

Ig gamma-1 chain C region - rat

C/Species: Rattus norvegicus (Norway rat)

C/Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 16-Jul-1999

C/Accession: PS0017, C25941

R.Brueggemann, M.

Gene 74, 473-482, 1988

A/Title: Evolution of the rat immunoglobulin gamma heavy-chain gene family.

A/Reference number: PS0017, MUID:89232738, PMID:3149946

A/Accession: PS0017

A/Molecule type: DNA

A/Residues: 1-326 <BRU>

R.Brueggemann, M.; Free, J.; Diamond, A.; Howard, J.; Cobbold, S.; Waldmann, H.

Proc. Natl. Acad. Sci. U.S.A. 83, 6075-6079, 1986

A/Title: Immunoglobulin heavy chain locus of the rat: striking homology to mouse antibody

A/Reference number: A25941, MUID:86287397, PMID:3016742

A/Accession: C25941

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 220-326 <BR2>

C/Genetics:

A/Introns: 98/1; 113/1; 220/1

C/Superfamily: Immunoglobulin C region; immunoglobulin homology

C/Keywords: Immunoglobulin

F/20-84/Domain: Immunoglobulin homology <IMN>

Query Match 24.3%; Score 828; Length 326;
Best Local Similarity 55.1%; Pred. No. 6.6e-42;

Matches 150; Conservative 48; Mismatches 48; Indels 26; Gaps 4;

QY 375 SQGVLLSNIKVLPTWSTVPEKSCDKTH-----TCPPCAPELLGG 416

Db 59 SGLVLTLSVTV-PSSTWPSQVTCNVAPASSTKVDKXIVRNCGDCKPC---ICRG 113

QY 417 ---PSFLPPPKKDTLMSRTPEVTCVVVDVSHEDPEVKFNKYVDGVENNAKTRPEE 473

Db 114 SEVSVFIFPPPKDVLITTLTPKVCVVVDISQDDPEVNFVFDVDEVHTAQTPEPE 173

QY 474 QVNSTRVVSVLTVLHODMNGEKYCKSNKALPAIEKTIISKAGQREPOVYTLPPS 533

Db 174 QFVNSTRVSVSELTFLHODMNGRTPEKCTSAAPSPIEKTIISKPEKRTQVPPVYTWSPF 233

QY 534 RDELTKQVSLTCLVKGFPSPDIAVWESNGQPENNYKTPPYLDSGSEFFLYSKLTVDK 593

Db 234 KEEMTQNEVSITCMVAGFPDPDIYVEMQNGQPENNYKTPPYLDSGSEFFLYSKLTVDK 293

QY 594 SRWQGNVPSGSVMHEALNNHYTQKLSLSPG 625

Db 294 EKWQGNVPSGSVMHEALNNHYTQKLSLSPG 325

RESULT 36

GLMS

Ig gamma-1 chain C region, secreted form - mouse

C/Species: Mus musculus (house mouse)

C/Date: 31-Mar-1980 #sequence_revision 24-Sep-1981 #text_change 16-Jul-1999

C/Accession: A02159, A26234, A26236

R.Hono, T.; Obata, M.; Yamawaki-Kataoka, Y.; Kataoka, T.; Kawakami, T.; Takahashi, N.;

Cell 18, 559-568, 1979

A/Title: Cloning and complete nucleotide sequence of mouse immunoglobulin gamma1 chain g

A/Reference number: A02159, MUID:80045036, PMID:115593

A/Accession: A02159

A/Molecule type: DNA

A/Residues: 1-324 <HON>

A/Cross-references: GB:J00453

A/Note: the sequence was determined from the germ-line gene

A/Note: Lys-324 is removed posttranslationally

R.Obata, M.; Yamawaki-Kataoka, Y.; Takahashi, N.; Kataoka, T.; Shimizu, A.; Mano, Y.; Se

Gene 9, 87-97, 1980

A/Title: Immunoglobulin gamma-1 heavy chain gene: structural gene sequences cloned in a

A/Reference number: A26234; MUID:80202559; PMID:6769752

A/Contents: MOPC 31C

A/Accession: A26234

A/Molecule type: mRNA

A/Residues: 76-324 <OBA>

A/Cross-references: GB:V00775; NID:951652; PIDN:CAA24153.1; PID:951653

R.Rogers, J.; Clarke, P.; Salsner, W

Nucleic Acids Res. 6, 3305-3321, 1979

A/Title: Sequence analysis of cloned cDNA encoding part of an immunoglobulin heavy chain

A/Reference number: A26236; MUID:80012837; PMID:113776

A/Contents: MOPC 21

A/Accession: A26236

A/Molecule type: mRNA

A/Residues: 170-275, 'D', 277, 'D', 279-322 <ROG>

A/Cross-references: GB:V00795; NID:951830; PIDN:CAA24176.1; PID:9780265

R.Aderugbo, K.

J. Biol. Chem. 253, 6068-6075, 1978

A/Title: Evolution of immunoglobulin subclasses. Primary structure of a murine myeloma g

A/Reference number: A26237; MUID:78242288; PMID:98524

A/Contents: annotation; MOPC 21

A/Note: this is the final paper in a series reporting the protein sequence, the disulfid

A/Note: there are a number of differences from the sequence shown

C/Genetics:

A/Introns: 1/1; 98/1; 111/1; 218/1

C/Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kap

hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la

C/Superfamily: Immunoglobulin C region; immunoglobulin homology

C/Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglob

F/20-84/Domain: immunoglobulin homology <IM1>

F/98-110/Region: hinge

F/131-200/Domain: immunoglobulin homology <IM2>

F/237-304/Domain: immunoglobulin homology <IM3>

F/27-92, 138-198, 244-302/Disulfide bonds: #status experimental

F/102/Disulfide bonds: interchain (to light chain) #status experimental

F/104,107,109/Disulfide bonds: interchain (to heavy chain) #status experimental

F/174/Binding site: carbohydrate (asn) (covalent) #status experimental

Query Match 24.2%; Score 826.5; DB 1; Length 324;
Best Local Similarity 60.0%; Pred. No. 8e-42;

Matches 144; Conservative 46; Mismatches 45; Indels 5; Gaps 2;

QY 388 PWSFVPEPKSDKNTKRP--CPARELLGSGSVLFPPKPKDTLMSRTPEVTCVVVDV 445

Db 87 PWSFVPEPKSDKNTKRP--CPARELLGSGSVLFPPKPKDTLMSRTPEVTCVVVDV 143

QY 446 SHEDEVEKFNMYVDDEVENNAKTRPEEQYNSTVYVSVLTVLHODMNGEKYCKVSNK 505

Db 144 SKDDPEVQPSWVDDEVENHTAQTPEEQYFNSTPFSVSELPIMQDMVNGKPFKRVNSA 203

QY 506 ALPAPIETISKAKQPREPOVYTLPPSRDELTKQVSLTCLVKGFPSPDIAVWESNGQ 565

Db 204 AFPAPIETISKAKQPREPOVYTLPPSRDELTKQVSLTCLVKGFPSPDIAVWESNGQ 263

QY 566 PENNYKTPPYLDSGSEFFLYSKLTVDSKRNQGVPSGSVMHEALNNHYTQKLSLSPG 625

Db 264 PAENYKTPPYLDSGSEFFLYSKLTVDSKRNQGVPSGSVMHEALNNHYTQKLSLSPG 323

RESULT 37

S00847

Ig gamma-2c chain C region - rat (fragment)

C/Species: Rattus norvegicus (Norway rat)

C/Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 23-Jul-1999

C/Accession: S00847

R.Brueggemann, M.; Delmaestro-Galfre, P.; Waldmann, H.; Calabi, F.

Eur. J. Immunol. 18, 317-319, 1988

A/Title: Sequence of a rat immunoglobulin gamma-2c heavy chain constant region cDNA: ext

A/Reference number: S00847; MUID:8816903; PMID:3127222

A/Accession: S00847

A/Molecule type: mRNA

A/Residues: 1-329 <BRU>

A/Cross-references: EMBL:X07189; NID:957602; PIDN:CAA30169.1; PID:g663228

F:98-113/Region: hinge
F:137-206/Domain: immunoglobulin homology <IM2>
F:243-310/Domain: immunoglobulin homology <IM3>
F:15/Disulfide bonds: interchain (to light chain) #status experimental
F:127-82,144-204,250-308/Disulfide bonds: #status experimental
F:107,110,112/Disulfide bonds: interchain (to heavy chain) #status experimental
F:180/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 23.9% Score 814.5; DB 1; Length 330;
Best Local Similarity 47.4%; Pred. No. 4.2e-41;
Matches 165; Conservative 47; Mismatches 77; Indels 59; Gaps 7;

QY 285 PQAIPQVAGSGNLTALDAKTKGKHQEV-----NLVNNARATQLOKNTLCEVGTSPKRM 339
DB 34 PEPVLTWNSSGLSSGCVHTFPAVLOSDLYTLSSVTATVTSWTPSQSICNVNHPAS---- 89
QY 340 LSLKLENKAKYKSKKEKRVWVLNPEAGWQCILSDSGVLLSNIKVLPTWSTVEPKSC 399
DB 90 -----STVYDKKIEP-----RG-----PT----IKP--- 106
QY 400 DKHTTCP--CPAPBLGSPVFLPPPKDITLMSRTPEYTCVVDVSHDEPEKFNMY 457
DB 107 -----CPCKCKPAPVLGSPSVFIRPKIKOVLMSISPIYTCVVDVSEDDPPVQISWF 161
QY 458 VDGEVFNAAKTKPREEQVNSTYRVSVLTVLHOMLNGEKYKCVKSKALPAPLEKTIISK 517
DB 162 VNNVEVFAQTQTHREDNSTLRVVSALPIQHODWMSKEKFKCVKNNNDLPAPLERTISK 221
QY 518 AKGQREPOVYTLPPSRBELTKNOVSLTCLKGFPSPDIAVEMWNSGQENNYKTPPVYL 577
DB 222 PKGSFRAPQVYVLPPEBEEMTKQVTLTCMTDFMEDIIYEMWTNGKTELNYKNTPEVL 281
QY 578 DSDGSFPLYSKLTVDKSRMOQGNVSCSVMEALHNHYTOKSLSPG 625
DB 282 DSDGSFMYSKLRYEKKMVERNSYSCSVMEGLNHHHTTSFSPRTPG 329

RESULT 40
GZMSJ1
Ig gamma-2b chain - mouse
CISpecies: Mus musculus (house mouse)
CIDate: 31-Mar-1980 #sequence_revision 01-Dec-2000 #text_change 01-Dec-2000
CAccession: S25057; A02157; A26235; A26233; A53558
RFlischer, R.; Voess, A.; Niersbach, M.; Munzler, W.; Hirsch, H.J.; Kreuzaler, F.
submitted to the EMBL Data Library, July 1992
A:Description: Production of a Tobacco mosaic virus (TMV) inactivating neocort specific m
A:Reference number: S25057
A:Accession: S25057
A:Status: Preliminary
A:Species: Mus musculus
A:Molecule type: mRNA
A:Residues: 1-474 <FIS>
A:Cross-references: EMBL:X67210; NID:G54826; PION:CAA47649.1; PID:G54827
Riyamawaki-Katooka, Y.; Katooka, T.; Takahashi, N.; Obata, M.; Honjo, T.
Nucleur 283, 786-789, 1980
A>Title: Complete nucleotide sequence of immunoglobulin gamma2b chain gene cloned from r
A:Reference number: A02157; MUID:80120716; PMID:6766534
A:Accession: A02157
A:Status: a allele
A:Accession: A02157
A:Molecule type: DNA
A:Residues: 138-161, 'U', 163-189, 'FP', 193-474 <YAN>
A:Cross-references: GB:J00461
A>Note: The sequence was determined from the germline gene
R Tucker, P.W.; Marcu, K.B.; Slightom, J.L.; Blattner, F.R.
Science 206, 1299-1303, 1979
A>Title: Structure of the constant and 3' untranslated regions of the murine gamma2b hee
A:Reference number: A26235; MUID:80081501; PMID:117548
A:Accession: MPC 11
A:Content: A26235
A:Accession: A26235
A:Molecule type: mRNA
A:Residues: 138-172, 'P', 174-189, 'FP', 193-376, 'T', 378-474 <TUI>
A>Note: Lys-474 is probably removed posttranslationally
R Tucker, P.W.; Marcu, K.B.; Newell, N.; Richards, J.; Blattner, F.R.
Science 206, 1303-1306, 1979

A>Title: Sequence of the cloned gene for the constant region of murine gamma2b immunoglobulin A.
A.Reference number: A26232; MUID:80081502; PMID:117549
A.Accession: A26232
A.Molecule type: DNA
A.Residues: 138-172, 'P', 174-189, 'FP', 193-376, 'T', 378-474 <TU2>
R.Oligo: R.; Rougeon, F.
Nature 296, 761-763, 1992
A>Title: Mouse immunoglobulin allotypes: post-duplication divergence of gamma2a and gamma2b alleles.
A.Reference number: A26233; MUID:82173203; PMID:6803173
A.Contents: b allele
A.Accession: A26233
A.Molecule type: DNA
A.Residues: 138-161, 'L', 163-189, 'EP', 193-300, 'R', 302-331, 'A', 333-437, 'DI', 440-474
A.Cross-references: GB:U00461
R.Kim, H.; Yamaguchi, Y.; Masuda, K.; Matsunaga, C.; Yamamoto, K.; Irimura, T.; Takahashi, J. Biol. Chem. 269, 12345-12350, 1994
A>Title: O-glycosylation in hinge region of mouse immunoglobulin G2b.
A.Reference number: A53598; MUID:94216359; PMID:7512967
A.Accession: A53598
A.Status: Preliminary
A.Molecule type: protein
A.Residues: 234-251 <IM>
C.Comment: The a allele sequence is shown.
C.Genetics:
A.Introns: 138/1; 236/1; 258/1; 368/1
C.Complex: An immunoglobulin heterotrimer subunit consists of two identical light (kappa) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a Cyspeptide family: immunoglobulin C region; immunoglobulin homology
C.Keywords: alternative splicing; duplication; glycoprotein; heterotrimer; immunoglobulin F157-222/Domains: immunoglobulin homology <IM1>
F1236-257/Region: hinge
F1281-350/Domains: immunoglobulin homology <IM2>
F1387-454/Domains: immunoglobulin homology <IM3>
F1157/Disulfide bonds: interchain (to light chain) #status predicted
F1164-220,288-348,394-452/Disulfide bonds: #status predicted
F1247,250,253,256/Disulfide bonds: interchain (to heavy chain) #status predicted
F1324/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 23.8%; Score 812; DB 1; Length 474;
Best Local Similarity 33.7%; Pred. No. 9,4e-41;
Matches 203; Conservative 72; Mismatches 154; Indels 174; Gaps 15;

Dy 34 GDVLTCTASQKKSIOF--HWKNSNQIKILNQG---SFL--TKGPSKLNDADRSRRS 85
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 34 GASVKMSCKASGVFIYVWWM-----VKQKPGCGLEWIGVINPNKDGTGFENKFPGKAT 88

Dy 86 LW---DGANPELLIKLKIEDSDTYICEVDQKEBVOVLVGLTANSPTHLGOSLTLT 142
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 89 LTPDSKSNNTAYMEISSLTSBDSAIVYYCARD-----YDWPFAYWGQETLVLT 134

Dy 143 LESDPGSSPSVQCRSPKGNIOGGKTLVSQLELDQSGTWTCVTVLNQKKVEFKIDIVL 202
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 135 VSAKTTPEPVYPLAPCCGDTTGSSTVS----- 162

Dy 203 AFQKASSIVYKKEGEQVEFSFPLAFTEVKLTGSGELMWQAERASSSKSWITFDLNKKEVS 262
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 163 -----GCIVKGYPESVTVT-----MNNGSLSS----- 186

Dy 263 VKRVTDPPKLMGKKLPRLHLLTPALFOYAGSGNLTALEAKTGKLGHEYNLVMPARQL 322
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 187 -----VH--TLSQLAQ---SGLYTMS-----SSVVPSPSTPV 214

Dy 323 QKNLTCTCWGPTSPKMLSLKLENKAAXYSKRERPAWVNLEPAGMMOCLLSDSGOVLLES 382
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 215 SQYTCTGVAAHPAS-STTVDDKLL-----PGCPSTTNV----- 246

Dy 383 NIKVLPWTSTPVEBKSCDKTHTCPCPAPELLGSPSVLFPRPKXDTLMISRTPEVTCVV 442
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 247 -----CRPCECHKCAPRLLEGSPVFIPRPNIKDVLMISLRPKVTCVV 290

Dy 443 VDSSHEDPEYKFNMYVDGVEVNAKTFRREOYNSTRVVSULTVLHOMLNGEKYKCY 502
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 291 VDSEDDPDVQISFNVNNVEHTAOOTRHEDNINSTRIVSTPLRIHOHMSGEEKYKCY 350

[illegible][illegible]


```

Oy 209 SILVKKKEBOUEFS-----PFLAFTVEKUTS-----GELMWOAERSS 247
Db 130 DIDLMGHTEVAVSSGPTAPSVFPISSCSGSTQOQPVVGLCATGYIFGPVFSKSGASG 189
Oy 248 SKSWITPDLKKEKESVKREVTODPKLQMGKKULPLHLTPQALPOVAGSGNL---TLALBAK 304
Db 190 ATS-----VTVPETHGVGP---HKRASFLRPHAGADPFPGSVNHQAT 230
Oy 305 TGKJHQEVLVYMR-----TQLOKULTCIEWGPTSKMLISLKENKEAKVSKREKPV 358
Db 231 RTSLOVNEGCAGAGEPTPPEVOVLHSSVCSTLGDSSVELLCVI-----TGFSPPPV 282
Oy 359 ---WVLNBEAGMOCILSDSGOVLLSNIKV-LPTMSTPV--EPKSCDKTHC-----PP- 407
Db 283 EVELMLVD---GAPALVATMTMRPQREAGSKTYMALISQVNRREDMKAKAKATCRKMHQAT 339
Oy 408 -----CPABELLQGPS-----VFLPFPKPXDTLMISRTPEVTCVVVDVSHEDDEV 452
Db 340 GGTAGOHARFCGSGS---GAOSCSPIQIFVVPSP-P-GSLYIRIDAKVHCLVYNLP-SDAISL 394
Oy 453 KFNWVVDVEVHNNAKTKPREQYNSTYRVVSVLTVLHDQMLNGEKYKCKVSNKLLPAPIE 512
Db 395 SISWTRÉSGALRPDPMYLTHEFNQGTFTASSLSLAISTDQMDLAGEFTTVOGHEDLPVLG 454
Oy 513 KTSIKAKQPREPQVYTLPPSRDELTKNOVSLTCLVKGFSPSDIAVEN--ESNGCPENNY 570
Db 455 KSIAGHAGKVTAPAFYIFTPPHAEELSLAEVLTITGLVRGFEQEHVEVQLRRHNSVPAAEF 514
Oy 571 KTTPLVLD--SDGSFPLYSKLTVDKSRWQOGNVSCSMHEALNNHYTQKSLISLSPG 625
Db 515 VTPPLKKNQDGTGTFLLYSKMTVPASWQGVSAQWVHGGLPMRFQRPQLOKTPG 571

```

```

RESULT 50
SI4683
Ig mu chain precursor, membrane-bound (clone 201) - human
C|Species: Homo sapiens (man)
C|Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 23-Jul-1999
C|Accession: SI4683; 508047
R|Friedlander, R.M.; Nusensweig, M.C.; Leder, P.
Nucleic Acids Res. 18, 4278, 1990
A|Title: Complete nucleotide sequence of the membrane form of the human IgM heavy chain
A|Reference number: SI4683; MUID:90332450; PMID:2115996
A|Accession: SI4683
A|Molecule type: mRNA
A|Residues: 1-627 <FRI>
A|Cross-references: EMBL:X17115; NID:g33450; PIDD:CAA34971.1; PID:g33451
C|Superfamily: Immunoglobulin C region; Immunoglobulin homology
C|Keywords: Immunoglobulin; membrane protein
F|1-15/Domain: signal sequence #status predicted <SIG>
F|16-627/Product: Ig mu chain #status predicted <MAT>
F|34-117/Domain: Immunoglobulin homology <IMM>

Query Match      11.4%  Score 390.5;  DB 2;  Length 627;
      Local Similarity 22.6%;  Pred. No. 1.2e-15;
      Matches 157;  Conservative 103;  Mismatches 231;  Indels 205;  Gaps 30;

QY      32  KKGPVELTCTAS-----QKKSIDFHAKNSNQIKLGNQSPFLTKGSKIND 78
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      32  KPGSSVKVSKASGCTFSSVYAIISWVRQAPGQGLEWM-GGILPFGT-ANVAKQFGRVTI 89
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      79  RADSRRSAMDGNPFILIKNLKIEDSDTYICEVDQKEEVOLL--VFGLTANSDTHL-- 133
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      90  TADEST-----TAMELSLSLRSEDTAYYC-----AKTGILGSSGWTIPNSDYIYG 138
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      134 --LQGSLITLLESPGSSPS-----VQGR----- 156
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      139 MDVWGQGTIVVSSGSASAPTLFLVSCENSPDSTSSVAVGCLADFLPDSITFSWKYN 198
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      157 -----SPRG--KNIOGKTLISVSOLEOD-----SQT--WTCYVLQNGKQKPEKDIYVL 202
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      199 NSDISSTGSPSVLNGGKRYAAITSOVLLRPSKDVMOGSTDHVVCKQVHPNGNKKVPLFVI 258

```

Qy	203	AF--OKASSIYKXKGEQVESPFLAFIVEKLTSSGELMMQABASSK---SWIFPDLK	257
Db	259	AEIPEPVSVFVPPRDS--FFGNP-----RSKSLICGATGFSPPQIQVSW-----	301
Qy	258	NKEVSVKRVTOBPKLQMGKULPLHLTLPLQALPOYAGSGNLLALAEATGKLHOEVLVVM	317
Db	302	-----LREGQV-----GGCVTTIDQYQAEKXES-----	324
Qy	318	RATOLQALTCVWGGTSPFKMLSLKLENKAKYKSKKEPWWVLNPEAGMOCILSDSQ	377
Db	325	-----GPTTYKVTSTLIKESD-----WL---SQSMFTGRVDHRL	357
Qy	378	VLLE-SNIKVLPTMSPTVPEPKSCDKTHHCPCPAPELLGSPFLPPEPKQDTLMSRTP	436
Db	358	TFQQAASMCVVDQDTR-----VVAIRPS-FASIFLTKST	393
Qy	437	EYTCVVVDVSHEDPEVKFNMYVDQVFNAAKTREREQYNSYTRVSVLTFLVLDQWLNGK	466
Db	394	KLTCFLDITLTYYD-SWTSISMTKQNGEAVKTHNISESHNATPSAVGEAASICEDDMSGE	452
Qy	497	EYCKCKSNKALPAPIEKITSKAGQP-REPOVYTLPPSRDLT-KNGVSLTCLVKGFPYS	554
Db	453	RFCCTVTHHDLSPKQITISRPKGVALHRPVYLLLPPEEQNLNRESATITCLVTGSPA	512
Qy	555	DIIVEWESNGQP-ENNNYKTPPVVD--SDGSFFLYSKLTLYDKSRMOQGVFSCSVNHEA	610
Db	513	DYFVQMGMGQGPLSPKRYTTSAPMEPEAPAPRYFAHSLITLYSEEMNTGGEYTCVVAHHA	572
Qy	611	LNNHYTQKSLSPGLQDLDTCAEAQDELDGLMTT	646
Db	573	LPRRVTERIVDKS-----TEGEVSADEGGEFNNLWAT	603

```

RESULT 51
ig heavy chain precursor - African clawed frog (fragment)
S04845
C/Spectes: Xenopus laevis (African clawed frog)
C/Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 26-Aug-1999
JAccession: S04845; S05695
R/Amenya, C.T.; Haire, R.N.; Litman, G.W.
Nucleic Acids Res. 17, 5388, 1989
A>Title: Nucleotide sequence of a cDNA encoding a third distinct Xenopus immunoglobulin
A/Reference number: S04845; MUID:89345103; PMID:2503814
A/Accession: S04845
A/Molecule type: mRNA
A/Residues: 1-549 <AME>
A/Cross-references: EMBL:X15114
R/Litman, G.W.
submitted to the EMBL Data Library, April 1989
A/Reference number: S05695
A/Accession: S05695
A/Molecule type: mRNA
A/Residues: 'LC','J','3'-308,'H','310-549<LIT>'
A/Cross-references: EMBL:X15114; NID:G64799; PID:g763031
C/Superfamily: Immunoglobulin C region; Immunoglobulin homology
C/Keywords: glycoprotein; heterotrimer; immunoglobulin
F:26-109/Domains: immunoglobulin homology <IMM>
F:281,234/binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match          11.4%; Score 389; DB 2; Length 549;
Best Local Similarity 25.6%; Pred. No. 1.2e-15;
Matches 153; Conservative 82; Mismatches 211; Indels 152; Gaps 25;

Oy      110 EVENQKEVQL--LVFGILT-ANSDTHLLD--GQSILTTLESPPGSFSSVCRRSPRK--   161
       |::|::|::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      21 EIKPSGISIKLSCKTSTSYTFPTNIWIHIQQVPKGKLMIGRIVPGADPDYSSSYOGRCRH    80

Oy      162 ----NIGGGKITSLVSOLDSGTWCTCVLQNKKVEFKIDIVLARQKASSIYYKKEG     217
       |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      81 ISTDNPSTPFLQIANLNKVEDTAITYCA-----REGV                      112

Oy      218 QVERSF-----PLAFTVETKLTSGEGLMWQAERASSSSKSWITFDLKKNKEV        261

```

```

Db      113 GYEDFYWGQGMVTVTSATLHAPSVFPLRPGCGS-----SSDSHTIGLGSTGF 162
Qy      262 SVKRVTDOPKLOMGKKLPLHLTLPOALPO---YAGSGNLTALAKTKLHQEVNLV-VM 317
Db      163 LPAPV--DYKNNSSGITSGLKNFPAVLQOOSGLFASSSQLTFLSMWAKKSECEVNEKP 220
Qy      318 RATLOLQKULTCE-----VWGPTSPK-----LMSLKENKEAKVSKREKPVV 360
Db      221 TSTKTOKIECQDEDEPIEPTVEIILQGPCASSKVELLCLITGVAPSEIKYH-----WL 274
Qy      361 LNPBAGMMQCLSDSGQV-----LLESNIKYLPWSTPVPKSCDKTKHT 404
Db      275 LN-----GQVYNISPSNSKPCKEENGTFSSRSKV---SVPKEDMNSEDSYT 317
Qy      405 CPPC-PA-----PELLGSPSVFLFPFKPKDTLMSRTEPVYCVVDVSHEDPE 451
Db      318 CKNTHPAHHTKTEASTKKCDETAITPKYDVLPRSPKD-LVYTKAKVYCVISRAAMTD-D 375
Qy      452 VKFNN-VYDGYEVNNAKTKPREEQYNSTYRVVSVLTVLHQMVLNGKEYKCVSNKALPAP 510
Db      376 LTVQMSRSDGKALAFDSAP-EKAYDGFPTVKSTLKI SPGMENKQCNCKVHPDLRSP 434
Qy      511 IEKTSKAKGQREPOVYTLPRSPDELTKNOVSLCLYKGFPSDIAYEWESNG--OPEN 568
Db      435 IEKTSQKQDPEGTETTLTLPPSDDELNDFLSLCLMKNRPQDIYVFWKXDGVTLEED 494
Qy      569 NYKTPPYLSDSGSEFL-YSKLTVDSRWQGNVFCGVMEHALNHYTOKLSLSPG 625
Db      495 YMTTTPVLEEBEEBFISFKLTIRASDMRGATVSCI---AAINTISQADIKQNRG 548

```

RESULT 52

MHSM

IG mu chain C region, membrane-bound form - mouse

```

C:Species: Mus musculus (house mouse)
C>Date: 31-Oct-1980 #sequence_revision 30-Jun-1991 #text_change 22-Jun-1999
C:Accession: A02167; A37517; B02166
R:Early, P.; Rogers, J.; Davis, M.; Calame, K.; Bond, M.; Wall, R.; Hood, L.
Cell 20, 313-319, 1980
A:Title: Two mRNAs can be produced from a single immunoglobulin mu gene by alternative R
A:Reference number: A02167; MUID:80222874; PMID:6771020
A:Accession: A02167
A:Molecule type: DNA
A:Residues: 436-476 <BAR>
A:Cross-references: GB:V00816; GB:J00444; NID:952343; PIDN:CAA24197.1; PID:952344
R:Rogers, J.; Early, P.; Carter, C.; Calame, K.; Bond, M.; Hood, L.; Wall, R.
Cell 20, 303-312, 1980
A:Title: Two mRNAs with different 3' ends encode membrane-bound and secreted forms of in
A:Reference number: A37517; MUID:80222873; PMID:6771019
A:Contents: MOPC 104E
A:Accession: A37517
A:Molecule type: mRNA
A:Residues: 410-476 <ROG>
A:Cross-references: GB:V00821; NID:952355; PIDN:CAA4202.1; PID:9817972
R:Kawakami, T.; Takahashi, N.; Honjo, T.
Nucleic Acids Res. 8, 3933-3945, 1980
A:Title: Complete nucleotide sequence of mouse immunoglobulin mu gene and comparison with
A:Reference number: A02166; MUID:81076590; PMID:6255422
A:Accession: B02166
A:Molecule type: DNA
A:Residues: 1-435; GKPTLYNVSIMSDTGCTCY' <KAM>
C:Comment: The sequence of residues 1-409 was assumed to be identical with the correspond
C:Comment: During differentiation, B lymphocytes switch from expression of membrane-bound
B.
C:Complex: An immunoglobulin heterodimer subunit consists of two identical light (kappa)
chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 18
C:Superfamily: Immunoglobulin C region; immunoglobulin homology
C:Keywords: alternative splicing; duplication; glycoprotein; heterodimer; immunoglobu
F:21-91/Domain: immunoglobulin homology <IMM1>
F:129-201/Domain: immunoglobulin homology <IMM2>
F:239-307/Domain: immunoglobulin homology <IMM3>
F:346-417/Domain: immunoglobulin homology <IMM4>
F:436-476/Domain: carboxyl-terminal <CTS>

```

F:14/Disulfide bonds: interchain (to light chain) #status predicted
 F:28-89,136-199,246-305,353-415/Disulfide bonds: #status predicted
 F:46,211,243,258,281/Binding site: carbohydrate (asn) (covalent) #status predicted
 F:216/Disulfide bonds: interchain (to heavy chain) #status predicted
 F:293/Disulfide bonds: interchain (to mu chain in another subunit) #status predicted

Query Match 11.0%; Score 375.5; DB 1; Length 476;
 Best Local Similarity 22.9%; Pred. No. 6,4e-15;
 Matches 127; Conservative 83; Mismatches 188; Indels 157; Gaps 18;

```

Qy      144 ESPFGSSPVQGNRP-RGNIOGKTLISVQLELDGSGTTCYVLQOKKVEFKIDIVL 202
Db      3 QSFVNVFPLVSCBPSDKNLVAMGCLARDLFTSTSTFN----- 43
Qy      203 AFQASSIVYKKEGEQVEFSPFLAFVTEKLTGSGELMWQBARASSKSWITFDLKNKVS 262
Db      44 -----YQNTVEVQ-----GRTPTLRGKGYLA 68
Qy      263 VKRVTDOPKLOM-----GKULPLHLTLPOALPOYAGSGNLTALAKTKL 308
Db      69 TSGVLSPKSILREGSDVELVCKIHGGKRDHLVPI-----AAVEM 110
Qy      309 HQEVNLV-----MRATLOLQKULTCEVWGPTSPKMLSLKENKEAKVSKREKPVV 362
Db      111 NPVNVFVPPRDFGSPAPRKSCLICATNFTPKPIVSWLKQKLVESGFTTDPVTIEN 170
Qy      363 ---PEA-----GMWQCLSDSGQVLLSNIKVLPWSTPVPKSCDK 401
Db      171 KGSPQYVYKYSTITLISEIMLNLYTTCVDRHGLTFLEK-----NV 212
Qy      402 THTCPPRABELLGSPVFLFPFKPKDTLMSRTEPVYCVVDVSHEDPEVKNVYDGV 461
Db      213 SSTCAASPSYDIL---TFTIPSPFAD-IFLSKANLTCLVSNLATIVE-TLINSWASQSG 266
Qy      462 EVNNAKTKPREEQYNSTYRVVSVLTVLHQMVLNGKEYKCVSNKALPAPLEKTSKAKQ 521
Db      267 EPLETKIKIESHPNGTFSAKVAVSVCEVDMNRKEVCVTHRDLPSPQKFTSK--- 322
Qy      522 PRE-----POVYTLPRSPDELTKNOVSLCLYKGFPSDIAYEWESNGO--PENNYKTT 573
Db      323 PNEHNKRPAYVLLPRAREQNLRESATVCLYKGFSPALISYOMLORGQLQDEKVTYS 382
Qy      574 PVLVD--SDGSFPLYSKLTVDKSRWQGNVFCGVMEHALNHYTOKLSLSPGLQDET 631
Db      383 APMEPEPAPGFYFTHSILTVTEEMNSGFTYCVVGHIALPHLTVERTVKS-----TEG 437
Qy      632 CAAKQDELDGLWTT 646
Db      438 EVNAEEGFENLWTT 452

```

RESULT 53

S25705

IG mu chain - sheep

```

C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 23-Jul-1999
C:Accession: S25705
R:Patil, S.; Nau, F.
Mol. Immunol. 29, 829-836, 1992
A:Title: Isolation and sequence of a cDNA coding for the immunoglobulin mu chain of the
A:Reference number: S25705; MUID:92342148; PMID:1635560
A:Accession: S25705
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-592 <PAT>
A:Cross-references: EMBL:X5994; NID:g1269; PIDN:CAA42611.1; PID:g1270
C:Superfamily: Immunoglobulin C region; immunoglobulin homology
C:Keywords: immunoglobulin
F:483-554/Domain: immunoglobulin homology <IMM>

```

Query Match 10.9%; Score 371.5; DB 2; Length 592;
 Best Local Similarity 22.7%; Pred. No. 1.5e-14;
 Matches 156; Conservative 96; Mismatches 253; Indels 181; Gaps 29;

```

QY 6 PFRHLLVLIQI---ALLPAATGKNVYLGKKGDVTELTCTASQKSIQFH--WKNSNQIK 60
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 3 PLMTLFLVLSAPRGVLSQVQLQSSGSLVKPSETLITCTVSSSLTVAHVSHIRASQK 62
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 61 ILNGQGSFLTKGPKSLNDRADSRSL--WDQGNFPLIIKNKIEDSDTYICEVEDQKE 117
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 63 MPFMLGVEKGKGTYYNPALKSRSLIARDTSKQVSLSSMAIDDTAVYCC----- 114
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 118 VOLLVGLTANSTHLLQ-----GQSLITLT---ESPQSSPSVQC-RSPKSKIQ--- 164
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 115 -----ASAGAYFLADVINGRGLVTVSSSSSHPKVPFLVCSVSPSDENVTALG 166
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 165 -----GKTLISVSOL---ELQDSGTWCT---VLQNKQKVEFKIDIV 200
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 167 CLARADVNPVSVSMSKFNSTVSSERFWTFPEVLRLDGLMSASQVALHSSTTGTDGVL 226
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 201 VLAFO--KASSIYKKEGEQVEFSPLAFT---VEKLTGSGELMWQAERASSSKSWITF 254
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 227 VCEVQHPKGEDVGHKGVPREVEVLSPVSVFVPPCNSLSGNG-----NSKSSLIQCAT 279
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 255 DLKKEKSV-----KRVTDPPKIQMGKULPLHLTPQALPQYAGSGNLTALBATGKL 308
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 280 DFGPKQISLSWFRDGRKIVSD-----ISGQVETVQSSPTTYR 317
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 309 HQEVLVVMRATQLOKNTLCEVWGPTSPKMLSLKLENKAKVSKREKPVWVLNPRAGMW 368
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 318 AYSVLITTEHEKWSQSAVYTCV-----EHNKNT----- 345
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 369 QCLLSDSGQVLTESNIKVLPTWSTPVEPKSCDTHTCPPCAPELLGSPVFLPPPKPD 428
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 346 -----FQKNAS-----SSCDAT---PPSPPI-----GVFTIIPSPFAD 373
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 429 TLMISTPPEYTCVYVDVSHEDPEVKFN--YVDC--VEVNAKTKPEBQNTSYRVVSLV 485
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 374 -FELTSAKSLSCVLTMLASYD--GLNLSWHSQNGKALETHYV---FERHNDTFSSARGBA 427
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 486 TVLHODMLNGKEYCKVSNKALPAPIEKTSKAKG--QPREPOYYLTTPSARDEL-T-QNQS 543
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 428 SVCSSEMSGEETTCVAAHLDLPPEKSAISKPDVAMKPPSYVLPPTREQLSLSESAS 487
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 544 LTLGVKGFYPSDIAVEMESNGCP--ENNYKTPPVLDD--SDGSFFLYSKLTVDKSRMOOG 599
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 488 VTCLVVGAPADVFVQWMLQGEPRVAKSKYVTSPPAPEPDPSAFVHSLITLTVEEDMSKG 547
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 600 NVFSCSVMEALHNHTQKSLSLSPG 625
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 548 ETYTCVVGHEALPHMVTERTVDKSTG 573
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

RESULT 54
S38864
Ig epsilon chain C region - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 24-May-2001
C:Accession: S38864
R:Klpp, B.; Becker, W.; Schlaak, M.
submitted to the EMBL Data Library, November 1993
A:Description: Combination of a defined specificity and desired isotype by cloning of an
A:Reference number: S38864
A:Accession: S38864
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-548 <KIP>
A:Cross-references: EMBL:Z27397; NID:G416537; PIDN:CAA81788.1; PID:G940782
C:Superfamily: Immunoglobulin C region; immunoglobulin homology
F:J53-421/Domain: immunoglobulin homology <IMM>

Query Match 10.8%; Score 370; DB 2; Length 548;
Best Local Similarity 24.2%; Pred. No. 1.6e-14;
Matches 157; Conservative 108; Mismatches 217; Indels 166; Gaps 31;

QY 30 LGKGDVTELTCTAS--QKSIQFH-----KNSNQIKLNGQGSFLTKGPKSLN----- 77

```

```

Db 11 LKPGGSLKLSAAGLFRSSYGMWNRQIPDKLEWATISSGGTY--TYPDYSKGRFT 69
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 78 -DRASRSRLMDQGNFPLIIKNKIEDSDTYICEVEDQKEVQLLVFGLTANSTHLLQ 136
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 70 ISRDNAKNTLY-----LQMSLKSIEDTAMYYCAQGVSTMIRFAVW-----G 111
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 137 QSLTTLSPSSPSVQCSPRGKNIGQKTLISVQLELQDSG-----TWTCTVYLQ 188
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 112 QGTLTVSAGKTTPPSVPLAPG---SAAQNSWVTLGLCYKGFPEPVTYTMNGSL- 166
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 189 NQKVEFKIDIVLAFOKASSIYKKEGEQVEFSPLAFTVEKLTGSGEL-----WQAE 243
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 167 -----SSGVH-----TPVALQSLYLTSSSVTPSSTWSE 198
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 244 RASSKSWITPPLKNEVSKRVTDPPKIQMGKULPLHLTPQALPQYAGSGNLTALBA 303
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 199 TVTCN---VAHPASSTKYDKKIVPRD---CGCK-----PCIVP----- 230
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 304 KTGKLEHGVNLY-----VMRATQLOKNTLCEVWGPTSPKMLSLKLENKAKVSKR 354
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 231 -----EVSSVFITPPPKQDVLTRSTIQ--LYCFYGHILNDVSVMLMDRELTDTLA 281
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 355 EKPVWVLNPRAGMWQCLLSDSGQV--LTESNIKVLPTWSTPVEPKSCDTHTCPPCAPEL 413
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 282 QI---VLIKEBK---LSTGSKNLTGQMMSESTFCCKYTSQGVDLAHTRRCPDHE- 334
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 414 LGGSPVFLPPPKQDMLMISRTPEYTCVVDV--SHEDPEVKFN-----WYVDGV 461
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 335 PRGVITTYLPPSPPLD--LYQNGAPKLTCLVVDLESKNNVNTNQKTSVSASQWY---T 390
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 462 EVHNAKTKPREQVNSTRYVSVLTVLHODMLNGKEYCKVSNKALPAPIEKTSKAKQ 521
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 391 KHN-----NATTSITSLPVAKDWEGGYCYVDHPPEKPYVASTIKTPGQ 440
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 522 PREPOYYLTTPSARDELTKNQVSLTCLVKGFPYSDIAVEMESNGCPENN--YKTPPVLDS 579
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 441 RAPEVYVPPPEE--SEDKRLTLCLIQNFPFEDISVQWMLGDKGISNSQHSHTTTP-LKS 498
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 580 DGS---FLYSKLTVDKSRMOQNVFSCSVMEALHN--HYTQKSLSL 623
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 499 NSNNGFFIFSRLEVAKTLMTQRKQFTQVIEHALQKPRKLEKTIKSTS 546
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

RESULT 55
MHMS
Ig mu chain C region, secreted form - mouse
C:Species: Mus musculus (house mouse)
C:Date: 31-Oct-1980 #sequence_revision 31-Oct-1980 #text_change 16-Aug-1996
C:Accession: A02166; A26239; A26240; B02039
R:Kawakami, T.; Takahashi, N.; Honjo, T.
Nucleic Acids Res. 8, 3933-3945, 1980
A:Title: Complete nucleotide sequence of mouse immunoglobulin mu gene and comparison wit
A:Reference number: A02166; MUID:81076590; PMID:6255422
A:Accession: A02166
A:Molecule type: DNA
A:Residues: 1-455 <KAN>
A:Cross-references: GB:J00443
A:Note: the sequence was determined from the germline gene
R:Goldberg, G.I.; Vanin, E.F.; Zrolka, A.M.; Blatterner, F.R.
Gene 15, 33-42, 1981
A:Title: Sequence of the gene for the constant region of the mu chain of Balb/c mouse im
A:Reference number: A26239; MUID:82051295; PMID:6795090
A:Accession: A26239
A:Molecule type: DNA
A:Residues: 1-455 <GOL>
A:Note: the sequence was determined from the germline gene
R:Aufiray, C.; Rougeon, F.
Gene 12, 77-86, 1980
A:Title: Nucleotide sequence of a cloned cDNA corresponding to secreted mu chain of mous
A:Reference number: A26240; MUID:81165562; PMID:6260591
A:Contents: TPC183
A:Accession: A26240

```

A: Molecule type: mRNA
A: Residues: 1-725, 'N', 227-257, 'S', 259-367, 'K', 369-455 <AUP>
R: Kenny, M.; Stibley, C.; Fuhrman, J.; Schilling, J.; Hood, L.E.
Proc. Natl. Acad. Sci. U.S.A. 76, 2932-2936, 1979
A: Reference number: A6241; MUID:79223904; PMID:111247
A: Contents: annotation; MOPC 104E
A: Note: this sequence has been revised in reference A02039. Carbohydrate binding sites at
R: Kenny, M.R.; Fuhrman, J.S.; Schilling, J.W.; Rogers, J.; Stibley, C.H.; Hood, L.E.
Biochemistry 21, 5415-5424, 1982
A: Title: Complete amino acid sequence of a mouse mu chain: homology among heavy chain cc
A: Reference number: A02039; MUID:83075344; PMID:6815276
A: Contents: MOPC 104E

A:Residues: 177,'N','79-100','Q','102-225','N','227-257','T','259-367','K','369-455 <KEH>
C:Genetics: 1
A:introns: 1/1: 106/1: 219/1: 325/1
C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into the C:superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglob
F:21-91/Domain: immunoglobulin homology <IM1>
F:129-201/Domain: immunoglobulin homology <IM2>
F:239-307/Domain: immunoglobulin homology <IM3>
F:346-417/Domain: immunoglobulin homology <IM4>
F:436-455/Domain: carboxyl-terminal <CTS>
F:14/Disulfide bonds: interchain (to light chain) #status predicted
F:128-89/Disulfide bonds: #status experimental
F:46-21,249,281,442/Binding site: carbohydrate (Asp) (covalent) #status experimental
F:136-159,246-305,353-415/Disulfide bonds: #status predicted
F:216,454/Disulfide bonds: interchain (to heavy chain) #status predicted
F:293/Disulfide bonds: interchain (to mu chain in another subunit) #status predicted

[illegible]

```

RESULT 56
A24976
IG mu chain C region (allele b) - mouse
C|Species: Mus musculus (house mouse)
C|Date: 05-Jun-1988 #sequence_revision 05-Jun-1988 #text_change 21-Jan-2000
C|Accession: A24976
R|Schreier, P.H.; Quester, S.; Bothwell, A.
Nucleic Acids Res. 14, 2381-2389, 1986
A|Title: Allelic differences in mu-genes.
A|Reference number: A24976; MUID:66176735; PMID:3083402
A|Accession: A24976
A|Molecule type: mRNA
A|Residues: 1-455 <SCH>
A|Cross-references: GB:X03690; NID:G52381; PIDN:CNA27326.1; PID:G52382
A|Experimental source: strain C57BL/6
A|Note: the authors translated the codon AAG for residue 65 as Leu
C|Superfamily: immunoglobulin C region; immunoglobulin homology
C|Keywords: immunoglobulin
I|346-417/Domain: immunoglobulin homology <IMM>

```

Query Match	10.7%;	Score 364;	DB 2;	Length 455;
Best Local Similarity	22.8%;	Pred. No. 2.9e-14;		
Matches 122;	Conservative	80;	Mismatches 180;	Indels 152;
				Gaps 17;

QY	144	ESPSPSSPVQCSPP-RGKKIIOGCKTSLVSOLEIJDOSGTYCTVLQIOKQKVEFIDIVL	202
Db	3	QSPFNPPVLVSCSPSLSDKDLVAMGCLARFLLSTISIFTN-----	43
QY	203	AFQKASSIVYKKEGOVEFSFPFLAFVTEKLTGSGELMOMERASSKSMITFDLONKEVS	262
Db	44	-----YONNTEVIO-----GJTFPLTRGTGYULA	68
QY	263	VKRYTODPKLOM-----GKULPLMLTRPOLPOYAGSNGLLTALAEKTKGL	308
Db	69	TSQVLLSPKSLIEGSDYLVCKHIGKKNKDLHPRP-----AVAEH	110
QY	309	HOEVNLV-----MRATOLOKULTCESWGPSTPKMLSLKLENKEAKVSREKRPVVLN	362
Db	111	NPVNVNVPPRDPGFGSPAPRKSKLICGATNFTPKPIVSMWKQGLVESGTTDPVTIEN	170
QY	363	-----PEA-----GMQGLLSDSQGVLLSNIKVLPTWSPVPEKSCDK	401
Db	171	KGSTPQYKYKISTITLISEIMWLNLYTCVDBHGLTFLE-----NV	212
QY	402	THTCRPPCABELLGGPSVFLPRPKPKDMLISRTEPVTCVVDVSHEDPEVKENYVDGV	461
Db	213	SSITCAASPTDIL---TFITIPSPFAD-IFLSKANLTCLVANSIATYE-TLNTISWASQSG	266
QY	462	EYVNAKTRPEEDQYNSTYRVVSVLTLYHOMLWKEKCVKSVSKALPAPLEKITSKRAKQ	521
Db	267	EPLETKIKIMESHNGFTFSKAGVAASVCEBDMNNRKEVCVTHDPLSPQKGFISK---	322
QY	522	PRE-----POVYTLPPSRDEL-T-KNOVSLTCLVKGFPSPDIAYEWESNQ--PENNYKTT	573
Db	323	PNEVHKHPRVAVLLPRRREGDNLNRESATVTCVLKGFSPADISVOMLQRGDILLPQEKYVTS	382
QY	574	PPVULD--SDGSFPLXKLTLYDKSRMOQGNVFGSGVMEALAHNYTQKSLSSPSG	625
Db	383	APMPEBAPGFYTHSLITLVTEEBMNSGELYTTCVNGEALPHLVTERTVDKSTG	436

```

RESULT 57
EHMS
IG: epsilon chain C region (version 1) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 13-Jun-1993 #sequence_revision 13-Jun-1993 #text_change 16-Jul-1999
C:Accession: A02144
R:LiLu, F.T.; Albrandt, K.; Sutcliffe, J.G.; Katz, D.H.
Proc. Natl. Acad. Sci. U.S.A. 79, 7852-7856, 1982
A:Title: Cloning and nucleotide sequence of mouse immunoglobulin epsilon chain cDNA.
A:Reference number: A02144; MUID:83117774; PMID:681553
A:Accession: A02144

```

A: Molecule type: mRNA
A: Residues: 1-388 <Lit>
A: Cross-references: GS:J00476; NID:g194875; PID:AAA38085.1; PID:g387220
C: Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kg) chain disulfide bonds. In some cases, such as Iga and IgM, the subunits associate into 1a
C: Superfamily: immunoglobulin C region, immunoglobulin homology
C: Keywords: duplication; glycoprotein; heterotetramer; immunoglobulin
F: 1-44/Domain: immunoglobulin homology (fragment) <IM1>
F: 81-149/Domain: immunoglobulin homology <IM2>
F: 186-254/Domain: immunoglobulin homology <IM3>
F: 260-351/Domain: immunoglobulin homology <IM4>
F: 10,51,62,133,205,228,332,382/Binding site: carbohydrate (Asn) (covalent) #status predicted

A:Experimental source: cell line Ab 63

C:Genetics:

A:Map position: 14q32

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: glycoprotein; heterotrimer; immunoglobulin

F:121-90/Domain: immunoglobulin homology <IM1>

F:127-199/Domain: immunoglobulin homology <IM2>

F:237-305/Domain: immunoglobulin homology <IM3>

F:344-415/Domain: immunoglobulin homology <IM4>

F:14/Disulfide bonds: interchain (to light chain) #status predicted

F:28-88, 134-197, 244-303, 351-413/Disulfide bonds: #status predicted

F:46, 205, 272, 279, 440/Binding site: carbohydrate (Asn) (covariant) #status predicted

F:291, 452/Disulfide bonds: interchain (to heavy chain) #status predicted

F:291/Disulfide bonds: interchain (to mu chain in another subunit) #status predicted

Query Match 10.5%; Score 358.5; DB 2; Length 453;

Best Local Similarity 27.6%; Pred. No. 6,1e-14;

Matches 83; Conservative 62; Mismatches 11; Indels 45; Gaps 10;

QY 332 GPTSPKMLSLKLENKAKSKREKPVWLNPEAGMOCILSDSGQVLE-SNITKVLPTW 390

DB 172 GPTTYKYVSTLTIKESD-----WL-----SOSMFCRVDHGLTFQGNASSMCVDPDQ 218

QY 391 STPEPKSCDTHNCPCEPAPELLGSPVFLPPEPKDPTLMISRPETVCVVVSHEDP 450

DB 219 DTAIR-----VFAPPS-FASITLVSTKLTCLVTDLTYYD- 253

QY 451 EVKFNWYDGVENHAKTPREEQNSTYRVVSVLTVLHODMLNGEKYCKVSNKALPAP 510

DB 254 SVTISWTRQNEAKYKTHNISESHNPATFSAVGEASICEDEDMNGERFTCVHTDLPSP 313

QY 511 TEKTSKAKGP-REPOVYTLPPSRDEL-T-KNOVSLTCLVKGYPSDIAVESNGQP-- 566

DB 314 LKQTSRPEKGVALHRPDVYLLPPAREQLNRESATITCLVGFSPADVFWQMORQPLS 373

QY 567 ENNYKTPPVVD--SDGSFELYSKLTVDKXSGQGNVSSGVMEALHNHTQKSLSP 624

DB 374 PEKVYTSAPMEPEQAPGRYFAHSILTVSEEMNGETTYCVAAHEALPNRTERTVDKST 433

QY 625 G 625

DB 434 G 434

RESULT 64

MRBM

Ig mu chain C region, membrane-bound form - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 04-Dec-1986 #sequence_revision 30-Jun-1991 #text_change 23-Aug-1997

C:Accession: A02164

J:Bernstein, K.E.; Alexander, C.B.; Reddy, E.P.; Mage, R.G.

J:Immunol. 132, 490-495, 1984

A:Title: Complete sequence of a cloned cDNA encoding rabbit secreted mu-chain of V-Haz a

A:Reference number: A02164; MUID:84088930; PMID:6418803

A:Contents: a2 allotype

A:Accession: A02165

A:Molecule type: mRNA

A:Residues: 439-479 <BE2>

A:Accession: A02164

A:Molecule type: mRNA

A:Residues: 1-438, 'GKPTLVNLSIMSDTASTCY' <BER>

A:Note: The sequence of residues 1-438 was assumed to be identical with the corresponding

C:Complex: An immunoglobulin heterotrimer subunit consists of two identical light (kappa)

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: alternative splicing; duplication; glycoprotein; heterotrimer; immunoglobulin

F:121-92/Domain: immunoglobulin homology <IM1>

F:130-202/Domain: immunoglobulin homology <IM2>

F:242-310/Domain: immunoglobulin homology <IM3>

F:348-420/Domain: immunoglobulin homology <IM4>

F:439-479/Domain: carboxyl-terminal <CTS>

F:14/Disulfide bonds: interchain (to light chain) #status predicted

F:28-90, 137-200, 249-308, 356-418/Disulfide bonds: #status predicted

F.46; 114,212,261,277,284/Binding site: carboxylate (Aan) (covalent) #status predicted
F.219/Disulfide bonds: Interchain (to heavy chain) #status predicted
F.219/Disulfide bonds: Interchain (to mu chain in another subunit) #status predicted

```

Query Match          10.5%; Score 358.5; DB 1; Length 479;
Best Local Similarity 24.6%; Pred. No. 6,5e-14;
Matches 121; Conservative 83; Mismatches 175; Indels 113; Gaps 21;

QY 223 PPL-----AFVVEKLTGSGELMWQAEKASSKSTTFPLKAKVEVAVKVTQDPKLGK 277
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 9 YPLVSCGALTDGNLVAMGCLARDFLPSSVTFSW-SFK-NNSEISSRYKRTFPVVRGDK 66
QY 278 -----LPHLTLPLQAL-----POYA 292
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 67 YMATSQVLVPSKDVLOQTEETVLCVKQHSNNRDLKRSFPVDSLELPPNVSVTFIPRDSFS 126
QY 293 GSGNLTLLALEAK-TGKLHQEYNLVVMRATQ-----LQKNLTCEVGEPTSPKMLSLKLE 345
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 127 GSGTRKSLICQATGCFSPKQISVSWLRDQGVESGVLRKPVAEKLRKGAPATFSSMLT 186
QY 346 NKEAKVSKREKPVWVLNDEAGMOCCLSDSQVLLSNIKVLPTWSTPVEKSCDKTHTC 405
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 187 ITESD-----WL-----SQSLYTCRVDRHG-IFEDKAVSW-----SSSC 219
QY 406 PPCPAPELLGSPSVLPFPKPKDTLMISRTETVCVWVDVSHDEDEVKFNWVDVEVYN 465
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 220 STTSPS-----GIQVFPAPSPADT-FLSKSARLCLLVLDLTYYG-SLNISSW-----ASIN 268
QY 466 AKT-----KPREEQNSTYRVSVLTVLHOMLNKEKCYKSNKALPAPIETISKAG 520
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 269 GKALDTMHNITESHNNATFFSAMGEASVCAEDMESEQCTCTVTHADLPFLKHTISKRE 328
QY 521 QPRE-POVYTLTPSRDELTKNOVSLTCLVKGFPSPDIADVEMESNGQP-ENNYKTTTPV 576
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 329 VAKHPAPVYVLPAREQLVRESATVTCLVAGFSPADVFVQMGQRCGPLSDSKYVTSAPA 388
QY 577 LD--SDGSFPLXKLTIVKSRMQGNVSCSVMHGALNNHTQKSLSLSPGLQDETCAE 634
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 389 PEPOAPGLYFHTSTLTVEEDMNGETFTCVGHGALPHMWTERTVKS-----TEGEVG 443
QY 635 AODGELDLGAWTT 646
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 444 AEEBGFENLWTT 455

RESULT 65
S00390
Ig gamma chain (clone 36) - chicken (fragment)
N/Alternate names: Ig mu chain
C/Species: Gallus gallus (chicken)
C/Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 11-Jan-2000
C/Accession: S00390
R.Patvari, R.; Avizit, A.; Lentner, F.; Zliv, E.; Tel-Or, S.; Bursstein, Y.; Schechter, I.
EMBO J. 7, 739-744, 1988
A>Title: Chicken Immunoglobulin gamma-heavy chains: limited VH gene repertoire, combina
A/Reference number: S00390; MUID:86283642; PMID:3135182
A/Molecule type: mRNA
A/Residues: 1-504 <PAR>
A/Cross-references: EMBL:X07174
A/Note: This sequence was determined from the differentiated gene
C/Superfamily: Immunoglobulin C region; Immunoglobulin homology
C/Keywords: immunoglobulin

Query Match          10.5%; Score 358; DB 2; Length 504;
Best Local Similarity 23.3%; Pred. No. 7,5e-14;
Matches 146; Conservative 89; Mismatches 206; Indels 186; Gaps 28;

QY 59 IILIDNOSFLTKGSKINDRADSRSLMDQGNFPLIITKILKTIEDSDTYIC----- 109
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 3 VVAISSTGSSNYG-SAVKGRATISRDN-GGSTLRQLNNLRABDTGYTCARDLGYDL 60
QY 110 ---EVEDQKEVOLLVFGLTANSDTHLLOQGSLLTLESPPGSSSPSVQCRSPRGKNIQGG 166

```



```

Db      61 YAGQIDAMGHGTEIV---SSASPT-----SPRLYPLASCCSDSAVPVAVG 104
      167 KTLVSQG---LEQDSGTWCTVLOKQKVERKIDIVLAFQKASSIVYKKEGEVSEF 223
      105 CLTSPSSAGGSIWEGSG-----GTAAGVSGTPTKLSF 138
      224 PLAFYVEKLTGSGELWMAERASSSKSWTF-----DLKNKEYSVKRYVQDRLQMGKTL 278
      139 VRLSPBEK-----RSPFCSAAPGALAKKEVQYGRV--DP----- 172
      279 PLHLTLPLQALPOYAGSGNLTALAEATGKLHDEVNLVNMRATQLOKNTLCEVWG--PTSP 336
      173 -----VPVAPEVQ-----VLHASSCTPSGSEVELL-----CLVTGSPAPA 210
      337 KMLSLAKLENKAKVSKREKPPWVNLPEAGMMOCLISDSGQV-----LLESNIKVL--- 387
      211 EV-----EVLVDGVGGL--LVASQSPAVRSGSTVSLSRVNVSGTD 249
      388 -----PTWSTVPER--KSC-DKTHCTPCCPAPBELLGSPVFLPPKPKDTLM 431
      250 WREGKSYSCRVNHPAINTVVEHDVYKCPGAGSCSP-----QLYAIIPSPBE-LY 299
      432 ISRTPEVTCVVVDVSHEDPEVKFMYVVDGEVHNNAKTKPR--BEQYNSTYRVVSVLTVYL 489
      300 ISDAKRLCLVNLPL-SDSSLSVTWTR--KSGNLKRPDMVLQEHNGTYSASAVPVST 356
      490 QMNLNKEKFKCKVKSKALPAPIEKTISKAKGQPREPVYTLPPSRDELTKNOVSLTCLVK 549
      357 QMNLSEERFTCVQHHELPPLSKSYRYRTGPTPLIYFPAHPHEILSLISVTLSCVLR 416
      550 GFYPSDIAVEM--ESNGQPENNYKTPPVLD-----SDG-SPLFYSKLTVDKSRMQ 598
      417 GRPRDIERLRLDHAIVATERVTTAVLPERTANGAGGDDGDTFFVYSKMSVETAKNG 476
      599 GNVFSCSVNHEALHNHYTQKSLSLSPG 625
      477 GTVFACMAVHEALPMKRFQRTLQKQAG 503

RESULT 66
160082
C1:Species: Homo sapiens (man)
C1:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 23-Jul-1999
C1:Accession: I60082
R1:Zverev, V.V.; Sidorov, A.V.; Nedospasov, S.A.; Malushova, V.V.; Udalova, I.A.; Andzha
Vopr. Virusol. 40, 100-102, 1995
A1:Title: [Nucleotide sequence of two exons of the human T-lymphocyte CD4 receptor gene].
A1:Reference number: I60082; MUID:95407135; PMID:7676667
A1:Accession: I60082
A1:Status: preliminary; translated from GB/EMBL/DBJ
A1:Molecule type: mRNA
A1:Residues: 1-71 <RES>
A1:Cross-references: GB:S79267; NID:g1086922; PID:AB35273.1; PID:g1086923
A1:Genetics:
A1:Intron: 17/1
C1:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology

Query Match      10.5%; Score 357; DB 2; Length 71;
Best Local Similarity 98.6%; Pred. No. 6.9e-15;
Matches 70; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

      QY      1 NNRGVPRHLVLVQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIOFMKNSNQIK 60
      DB      1 NNRGVPRHLVLVQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIOFMKNSNQIK 60
      QY      61 ILGNQSGFLTK 71
      DB      61 ILGNQSGFLTK 71

RESULT 67

```

```

EHR1
Ig epsilon chain C region - rat
C1:Species: Rattus norvegicus (Norway rat)
C1:Date: 17-Dec-1982 #sequence_revision 17-Dec-1982 #text_change 16-Jul-1999
C1:Accession: A93442; A90937; A02143
R1:Hellman, L.; Pettersson, U.; Engstrom, A.; Karlsson, T.; Bennich, H.
Nucleic Acids Res. 10, 6041-6049, 1982
A1:Title: Structure and evolution of the heavy chain from rat immunoglobulin E.
A1:Reference number: A93442; MUID:83064537; PMID:6292865
A1:Accession: A93442
A1:Molecule type: mRNA
A1:Residues: 1-429 <HEL>
A1:Experimental source: strain LOU/c/MsJ, immunocytooma IR2
R1:Kindsvogel, W.R.; Reddy, E.P.; Moore, J.M.; Faust Jr., C.H.
DNA 1, 335-343, 1982
A1:Title: A cloned cDNA probe for rat immunoglobulin epsilon heavy chain: construction, i
A1:Reference number: A90937; MUID:83182019; PMID:6820340
A1:Contents: myeloma IR162
A1:Accession: A90937
A1:Molecule type: mRNA
A1:Residues: 'N', 169-307, 'L', 309-342 <KIN>
C1:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (Xp
hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la
C1:Superfamily: immunoglobulin C region; immunoglobulin homology
C1:Keywords: duplication; glycoprotein; heterotetramer; immunoglobulin
F1:19-80/Domain: immunoglobulin homology <IM1>
F1:118-166/Domain: immunoglobulin homology <IM2>
F1:223-291/Domain: immunoglobulin homology <IM3>
F1:327-398/Domain: immunoglobulin homology <IM4>
F1:462, 99, 170, 240, 265, 369, 419/Binding site: carbohydrate (Asn) (covalent) #status predicte

Query Match      10.4%; Score 355.5; DB 1; Length 429;
Best Local Similarity 27.8%; Pred. No. 8.5e-14;
Matches 101; Conservative 69; Mismatches 114; Indels 79; Gaps 16;

      QY      324 KNLTCSEVWGPTSPKLMLSKLENKEAKVSKREKPPWVNLPEAGM----- 367
      DB      74 KNFTCHV--THAPSTFVS-----DLTIARPNINIKPVDLLHSCDDNAFHTIQL 123
      368 -----WQCLLSD-----SQVLLESNIKVLPTWS--TPVEPKCDKT 402
      124 YCFVYGHIDNVSIHW--LMDRKIYETHAONVLIKEGKLASTSRINITQQQMSRST 181
      403 HTCP-----PCPAPBELLGSPVFLPPKPKDTLMISTPVTCTVVVDVSHED 449
      182 FTCKVTSQGENYWAHTRCSDE--PRGVITVLIIPSPLD--LYENGTPVLTCLVLDLSEEE 239
      450 PEVKFNMYVDGEVHNNAKTKPREBOYNSTYRVVSVLTVLDQWLNKGEKCKVSNKALPA 509
      240 -NITVTVWRREKKKSGSASQSTKHNATITSLIPDADQWLGEGYQCRVDPHPFPK 298
      510 PIEKTISKAKQPREPVYTL-PPSRDELTKNOVSLTCLVKGFPYSDIAVEM--ESNGQ 566
      299 PIVRSITKAPGRSAPEVYVFLPREEB--KDKRTLCTLIQNFPEPDISVQWLQDSKLI 356
      567 ENNYKTTTPVLDSDGS---FFLYSKLTLYDKSRMOQGNVFSSGVNHEALHN-HYQKSL 622
      DB      357 KSQHSHTTP-LKXGNSNORFFIFSRLEYTKALMTQKQCTGRVIAHEALREPKLERTISK 415
      623 SPG 625
      DB      416 SLG 418

RESULT 68
MHUB1
Ig mu heavy chain disease protein (Boc) - human
C1:Species: Homo sapiens (man)
C1:Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 16-Jul-1999
C1:Accession: A02163
R1:Barnikol-Matanabe, S.; Mihaesco, E.; Mihaesco, C.; Barnikol, H.U.; Hilschmann, N.
Hoppe-Seyler's Z. Physiol. Chem. 365, 105-118, 1984
A1:Title: The primary structure of mu-chain-disease protein BOR. Peculiar amino-acid sequ

```


R:Doibly, T.W., Devono, J.; Croce, C.M.
Proc. Natl. Acad. Sci. U.S.A. 77, 6027-6031, 1980
A>Title: Cloning and partial nucleotide sequence of human immunoglobulin mu chain cDNA
A:Accession: A26244
A:Molecule type: mRNA
A:Residues: 298-386/436-452 <DOL>
A:Cross-references: GB:J00257; NID:g185053; PIDN:AA53508.1; PID:g185056; GB:J00258; NID:J00259
R:Takehashi, N.; Nakai, S.; Honjo, T.
Nucleic Acids Res. 8, 5983-5991, 1980
A>Title: Cloning of human immunoglobulin mu gene and comparison with mouse mu gene.
A:Reference number: 137748; MUID:81124312; PMID:6450943
A:Accession: 137749
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 347-370 <TAK1>
A:Cross-references: EMBL:V00562; NID:g33448; PIDN:CAA23825.1; PID:g929649
A:Accession: 137750
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 433-452 <TAK2>
A:Cross-references: EMBL:V00563; NID:g33454; PIDN:CAA23826.1; PID:g925684
R:Milnesco, E.; Barnikol-Watanabe, S.; Barnikol, H.U.; Milnesco, C.; Hilschmann, N.
Eur. J. Biochem. 111, 275-286, 1980
A>Title: The primary structure of the constant part of mu-chain-disease protein BOT.
A:Reference number: A02162; MUID:81066716; PMID:6777162
A:Contents: Mu-chain-disease protein Bot and revisions to sequence of Gal
A:Accession: A02162
A:Molecule type: protein
A:Residues: 1-17, 'BPS', '22', 'T', '24-82', 'N', '84-90', 'Z', '92-93', 'B', '95', 'B', '97-144', 'E', '146-162', 'E', '163-164', 'E', '165-214', 'G', '216-262', 'D', '264-295', 'D', '297-414', 'A', '415-416', 'A', '417-418', 'A', '419-420', 'A', '421-422', 'A', '423-424', 'A', '425-426', 'A', '427-428', 'A', '429-430', 'A', '431-432', 'A', '433-434', 'A', '435-436', 'A', '437-438', 'A', '439-440', 'A', '441-442', 'A', '443-444', 'A', '445-446', 'A', '447-448', 'A', '449-450', 'A', '451-452', 'A', '453-454', 'A', '455-456', 'A', '457-458', 'A', '459-460', 'A', '461-462', 'A', '463-464', 'A', '465-466', 'A', '467-468', 'A', '469-470', 'A', '471-472', 'A', '473-474', 'A', '475-476', 'A', '477-478', 'A', '479-480', 'A', '481-482', 'A', '483-484', 'A', '485-486', 'A', '487-488', 'A', '489-490', 'A', '491-492', 'A', '493-494', 'A', '495-496', 'A', '497-498', 'A', '499-500', 'A', '501-502', 'A', '503-504', 'A', '505-506', 'A', '507-508', 'A', '509-510', 'A', '511-512', 'A', '513-514', 'A', '515-516', 'A', '517-518', 'A', '519-520', 'A', '521-522', 'A', '523-524', 'A', '525-526', 'A', '527-528', 'A', '529-530', 'A', '531-532', 'A', '533-534', 'A', '535-536', 'A', '537-538', 'A', '539-540', 'A', '541-542', 'A', '543-544', 'A', '545-546', 'A', '547-548', 'A', '549-550', 'A', '551-552', 'A', '553-554', 'A', '555-556', 'A', '557-558', 'A', '559-560', 'A', '561-562', 'A', '563-564', 'A', '565-566', 'A', '567-568', 'A', '569-570', 'A', '571-572', 'A', '573-574', 'A', '575-576', 'A', '577-578', 'A', '579-580', 'A', '581-582', 'A', '583-584', 'A', '585-586', 'A', '587-588', 'A', '589-590', 'A', '591-592', 'A', '593-594', 'A', '595-596', 'A', '597-598', 'A', '599-600', 'A', '601-602', 'A', '603-604', 'A', '605-606', 'A', '607-608', 'A', '609-610', 'A', '611-612', 'A', '613-614', 'A', '615-616', 'A', '617-618', 'A', '619-620', 'A', '621-622', 'A', '623-624', 'A', '625-626', 'A', '627-628', 'A', '629-630', 'A', '631-632', 'A', '633-634', 'A', '635-636', 'A', '637-638', 'A', '639-640', 'A', '641-642', 'A', '643-644', 'A', '645-646', 'A', '647-648', 'A', '649-650', 'A', '651-652', 'A', '653-654', 'A', '655-656', 'A', '657-658', 'A', '659-660', 'A', '661-662', 'A', '663-664', 'A', '665-666', 'A', '667-668', 'A', '669-670', 'A', '671-672', 'A', '673-674', 'A', '675-676', 'A', '677-678', 'A', '679-680', 'A', '681-682', 'A', '683-684', 'A', '685-686', 'A', '687-688', 'A', '689-690', 'A', '691-692', 'A', '693-694', 'A', '695-696', 'A', '697-698', 'A', '699-700', 'A', '701-702', 'A', '703-704', 'A', '705-706', 'A', '707-708', 'A', '709-710', 'A', '711-712', 'A', '713-714', 'A', '715-716', 'A', '717-718', 'A', '719-720', 'A', '721-722', 'A', '723-724', 'A', '725-726', 'A', '727-728', 'A', '729-730', 'A', '731-732', 'A', '733-734', 'A', '735-736', 'A', '737-738', 'A', '739-740', 'A', '741-742', 'A', '743-744', 'A', '745-746', 'A', '747-748', 'A', '749-750', 'A', '751-752', 'A', '753-754', 'A', '755-756', 'A', '757-758', 'A', '759-760', 'A', '761-762', 'A', '763-764', 'A', '765-766', 'A', '767-768', 'A', '769-770', 'A', '771-772', 'A', '773-774', 'A', '775-776', 'A', '777-778', 'A', '779-780', 'A', '781-782', 'A', '783-784', 'A', '785-786', 'A', '787-788', 'A', '789-790', 'A', '791-792', 'A', '793-794', 'A', '795-796', 'A', '797-798', 'A', '799-800', 'A', '801-802', 'A', '803-804', 'A', '805-806', 'A', '807-808', 'A', '809-810', 'A', '811-812', 'A', '813-814', 'A', '815-816', 'A', '817-818', 'A', '819-820', 'A', '821-822', 'A', '823-824', 'A', '825-826', 'A', '827-828', 'A', '829-830', 'A', '831-832', 'A', '833-834', 'A', '835-836', 'A', '837-838', 'A', '839-840', 'A', '841-842', 'A', '843-844', 'A', '845-846', 'A', '847-848', 'A', '849-850', 'A', '851-852', 'A', '853-854', 'A', '855-856', 'A', '857-858', 'A', '859-860', 'A', '861-862', 'A', '863-864', 'A', '865-866', 'A', '867-868', 'A', '869-870', 'A', '871-872', 'A', '873-874', 'A', '875-876', 'A', '877-878', 'A', '879-880', 'A', '881-882', 'A', '883-884', 'A', '885-886', 'A', '887-888', 'A', '889-890', 'A', '891-892', 'A', '893-894', 'A', '895-896', 'A', '897-898', 'A', '899-900', 'A', '901-902', 'A', '903-904', 'A', '905-906', 'A', '907-908', 'A', '909-910', 'A', '911-912', 'A', '913-914', 'A', '915-916', 'A', '917-918', 'A', '919-920', 'A', '921-922', 'A', '923-924', 'A', '925-926', 'A', '927-928', 'A', '929-930', 'A', '931-932', 'A', '933-934', 'A', '935-936', 'A', '937-938', 'A', '939-940', 'A', '941-942', 'A', '943-944', 'A', '945-946', 'A', '947-948', 'A', '949-950', 'A', '951-952', 'A', '953-954', 'A', '955-956', 'A', '957-958', 'A', '959-96

OY		332	GPTSPKMLSLKLNNKAQVSRKREKVPWVLNPNBAGMOCLLSDSGVILE-SNTKVLPTW	390
Dd		172	GPTTKAVSTLTIKESD-----WL-----GQSFTCVDRGLTFQGNASSMCVPDQ	218
OY		391	STPVEPKSCDKHTCTPCPCAPELLGGPSVFLEPPPKRDTLMISRTPEVTCCVVVDVSHDP	450
Dd		219	DTAIR-----VFALPPS-FASIFLTKSTKLTLVTDLTITVD-	253
OY		451	EKKFMWYDVGVHNAKTFRREQNVSTRVSVLVTHDDYDLNGEKYCKVKSKALPAP	510
Dd		254	SVTISMTRONGAVAKTHNISSHPNATFSAGEAASICEDNMNSGERFCTVTHTDLPSP	313
OY		511	IETKSIAKAGCP-REPQVYTLTPPSRDELTKNOVSITCLVKGFPSPDIAYEBSNGCP--	566
Dd		314	LKOTISRPRGVALNHRPDVLLPREARQLNLRBATITTCIVTGSPRALVPQMQRGPDS	373
OY		567	ENNYKTPPEVLD-SDGSFFLYSKLTVDKSRMOQGVNFCSVNHEALHNHYTKSLISP	624
Dd		374	PEKYTSAPMEPPOAFGRYFAHSILTVSEEWMTGETYTE-VAHMALPNRYERTVDKST	432
OY		625	G	625
Dd		433	G	433

RESULT 73
S03961

Ig mu chain C region - house shrew (fragment)
C:Species: Suncus murinus (house shrew)
C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 21-Jan-2000
C:Accession: S03961
R.Ishiguro, H.; Ichihara, Y.; Namikawa, T.; Nagatsu, T.; Kurosawa, Y.
FEBS Lett. 247, 317-322, 1989
A>Title: Nucleotide sequence of Suncus murinus immunoglobulin mu gene and comparison with
A:Reference number: S03961; MUID:89232144; PMID:2497033
A:Accession: S03961
A:Molecule type: DNA
A:Residues: 1-457 <ISH>
A:Cross-references: EMBL:X13920
C:Genetics:
A:Introns: 106/1; 221/1; 327/1
C:Superfamily: Immunoglobulin C region; immunoglobulin homology
C:Keywords: immunoglobulin
F,241-309/Domains: immunoglobulin homology <IMM>

Query Match	10.2%;	Score 348;	DB 2;	Length 457;
Best Local Similarity	23.7%;	Pred.No.2.6e-13;		
Matches	109;	Conservative 90;	Mismatches 160;	Indels 100; Gaps 18;

OY		204	FQKASSIYKKEGEQVERSFPLAFTEKLTGSGELMWQAER-ASSSKSWITFDLN----	258
Dd		43	FNSSSI-----SQNIYNFPEVFTGKYMATSQVLLPSTAILOSTDYITCHKHTTGE	97
OY		259	KEVSYKRVTDQPKLQMGKLLPLHLT--POLPYAGSGNLTALEAKTGKLDHOENVLV	316
Dd		98	KEKVY-----ELQYTPELPNVNSIFVP---PRNSFSGN-----HPRTSOLI	135
OY		317	MRAITOLQKNLTCEVM-----GPTSPKMLSLKLNNKAQVSR	352
Dd		136	CQASGPSPTIWMJORCEPVQPSLVTSVAEEPRKSGPTTFVRVLSRLTTENE----	191
OY		353	KREKFWVVLNPBAGMOCLLSDSGVILLNESNIKVLPTWSSTPVEPKSCDKHTCTPCCAPPE	412
Dd		192	-----WLSQRE-----FTQCALHKG-LTFQKNVSSV-----CMGDD	221
OY		413	LIIGPSVFLPPPKPYDTLMISRTPEVTCVYVDVSHEDPEVKNYVDGVEVHNAKTGRE	472
Dd		222	TSTGISVFLPPTPAN-IFLTOSAOJLCIVTGATYVD-SIDISRSRNGEALQTHVNISE	279
OY		473	EOYNSITVAVSLVTLHDQWLNKGKEYCKVSNKALAPARIKTSIAKAGPRE-PQVYTLF	531
Dd		280	SHPNSTPAKGHASVCREEWSEGKFTCVQHSDLSPLKQSIJSRPDVANDPDSVFLP	339


```

RESULT 79
S60266
novel antigen receptor precursor - nurse shark
C:Species: Ginglymnotoma cirratum (nurse shark)
C:Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 21-Jul-2000
C:Accession: S60266
R:Greenberg, A.S.; Avila, D.; Hughes, M.; Hughes, A.; McKlincy, E.C.; Flajnik, M.F.
Nature 374, 168-173, 1995
A:Title: A new antigen receptor gene family that undergoes rearrangement and extensive
A:Reference number: S60266; MID:95183140; PMID:7877689
A:Accession: S60266
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-684 <GRE>
A:Cross-references: EMBL:U08701; MID:g659442; PID:AA048195.1; PID:g659443

Query Match
Best Local Similarity 23.5%; Score 330.5; DB 2; Length 684;
Matches 158; Conservative 83; Mismatches 237; Indels 195; Gaps 29;

QY 16 LALLPAAQGNKVVLOKKDVTVELTCTAS--QKKSIFHW-KNSQIKILGNSGFLTKG 72
DB 145 VLLLSHATEQRA-----NRVQLVCLISGYYPENIAVSMQNTKTI-----TSGFATTS 194
QY 73 PSKLANDRADSRRSL-----WDQGNFPLIKILKIEDSDTYICEVEDQKEEYQLLVFGL 125
DB 195 PVKTSNDSFCASLKVPIQEMSRGS-----VISCQV----- 226
QY 126 TANSDDTHLLQGSGLTLLTSPSSPSVQCRSRGNKIQGKTLVSQLELQDSGTWCT 185
DB 227 -SHSATSSNQKREIRST-----SEIAYLLRD-----TYEIIWIDKSAITLVE 268
QY 186 VIQONQKVFKIDIVLAQKASSIYKKEGEQVE--FSPLAFVKEKLTSGSELIMQA 242
DB 269 VISTVSAG-----VVVSMWVNGKV--RNEGVMETPKMSGNQYLISLTSSVEW--- 317
QY 243 ERASSKSWITFDLKNKEVSVKRVTDPKLQMGKPLHLTLPOALPOYAGSGNLTAL 302
DB 318 ---QSGVEYTCNAKQDQSTPVVKTAKARVEPTP-HRLLPSPSEIQTST----- 366
QY 303 AKTGKLQEVNLVWKRATQLOKNLTCGVWGPTSPKMLSLKLENKAKVSKREKPVWVLN 362
DB 367 -----SATLTCLIRGFYDPKSVS----- 385
QY 363 PRAGMWQCLLSDGVLLESNIKVLPTM-----STPEPKSCOKTHTC--- 405
DB 386 -----WQ-----KDDVSVSANVTNFPALQDLTFSTRSLNLITAEWMSGAK-YTCTAS 434
QY 406 -PP-----CPAPELLGSPVFLPPPKKDTLMISRTPEVTCVVVDVSHED 449
DB 435 HPSQSSTVKRVIRNOKVDCRQTDI-----SVSLKP-PFEIWTQQTATTICEIV--YSD 486
QY 450 PE-VKFNWYVDGEVHNNAKTPREQYNSTYRVSVLTVLHODWLNKGEYKCKVSNKALP 508
DB 487 LENIKVFQVNGVERKKGVEQTONPEWSSGSKTIVSLKXMASBEMDSGTETVCLVEDSELP 546
QY 509 AIEKTIKSAK-GQREPOVYTLTPSRDL-TKNQVSLCTLYKGFPPSIADVEMSGNG- 565
DB 547 TVVKASIRKANYSQMHPKYLHPSTDIIDENSATLMCLATNFPABEIIYGVMMANDLT 606
QY 566 PENNYKTPPVLDSDGSPFLYSKLTVDKSRMOQGNVFCSVNHEALH-----NHYYQK 618
DB 607 LDGRTQVDSKSGSSFTDLRLTLAEWNSDTTISCLVGHPSLNDRLIRSTKSNCK 666
QY 619 SLSPGQLDET 631
DB 667 PSNVSVVLSDT 679

```

```

C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 20-Sep-1999
C:Accession: S31436
R:Fellah, J.S.; Miles, M.V.; Schwager, J.; Charlemagne, J.
submitted to the EMBL Data Library, November 1992
A:Description: cDNA sequence of Ambystoma mexicanum upslon heavy Igy chain.
A:Reference number: S31436
A:Accession: S31436
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-433 <FEL>
A:Cross-references: EMBL:X69492; MID:962420; PID:962421
C:Superfamily: immunoglobulin C region; immunoglobulin homology

Query Match
Best Local Similarity 23.6%; Score 313.5; DB 2; Length 433;
Matches 106; Conservative 64; Mismatches 162; Indels 117; Gaps 19;

QY 216 GEVFEFPPLAFTVEKLTGSGELMWQARASSKSS-----WIT---FDLKNE 260
DB 34 GEPVEVSKAGSTT---MPGKTFPAQQAQTALSTSSQIRIPAYEKGTNSVSCVKKRP 90
QY 261 VSV-----KRVTDPKLQMGKPLHLTLPOALPOYAGSGNLTALAEATGKLH 309
DB 91 TSTEIHKTITTSACKKATSKPSVQV-----LQSCADPDGNGSI----- 129
QY 310 QEVNLVWKRATQLOKNLTCGVWGPTSPKMLSLKLENKAKVSKREKPVWVLNPRAGMWQ 369
DB 130 -----ELVCLISGYTPDNIQVRWLVDNKAAPLQGTSP---PKG--- 165
QY 370 CLLSGQGVLESNIKVLPT-WST-----PVEPKSCDKHTGCPAPALGAP 417
DB 166 ---DQGTFFSTTSQINVTKSDMASGDKTCKVEHATSTSRADTTHNCADSGTFP---QF 219
QY 418 SVFLPPPKKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGEVHNNAKTPRE----- 472
DB 220 KVFLLAPKARD-LYIANQPVVLCIKTKMENSD-SLSVTM-----KREGPEBAVYS 269
QY 473 EGY---NSTYRVSVLTVLHODWLNKGEYKCKVSNKALPAPLEKTIKSAKQRPPOVYT 529
DB 270 EYIDSDGTFTAMSLNTLNKNEWEGDEFTCKVKKFDLPFLPSRSVSKPTGSAFTWYV 329
QY 530 LPPSRDELTN-NOVELTCLVKGFPSPDIADVEMSGNG--PENNYKTPPVLDSD-----GS 582
DB 330 FAPHEMELANDYFVSLTCLVKSFSPPDIYIQKQKGSYIPSDKYVSMERQEAAGTAGCT 389
QY 583 FFLYSKLTVDKSRMOQGNVFCSVNHEAL 611
DB 390 YFSYMLTIOKSDMDKRETFCTVAHSAV 418

RESULT 81
S21461
T-cell surface glycoprotein CD4 (allele 1) - pig (fragment)
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 20-Feb-1995 #sequence_revision 19-Apr-1996 #text_change 21-Jan-2000
C:Accession: I47131; S21461
R:Guecateson, K.; Germana, S.; Sundt, T.M.
J. Immunol. 151, 1365-1370, 1993
A:Title: Extensive allelic polymorphism in the CDR2-like region of the miniature swine C
A:Reference number: I47131; MID:93329116; PMID:8335933
A:Accession: I47131
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-99 <GUG>
A:Cross-references: EMBL:X65629; MID:91928; PID:CAA46583.1; PID:9388232
C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
C:Keywords: glycoprotein; T-cell
F:3-81/Domain: immunoglobulin homology <IMM>

Query Match
Best Local Similarity 60.2%; Score 305.5; DB 2; Length 99;
Matches 59; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

```

Qy 32 KKGDVELTCTASQKSIQFMKNSNQTILGNGSFL-TKSPKLNDRADSRSLMDQG 90
 Db 1 KAGDLAEIPCHSSQKKNLPFWNKSNQTKILGHSFWHTASVIELTSLDSKKNMWDHG 60
 Qy 91 NPLIIRKLKIEDSTYICEVEDCKEEVOLVFGLTAN 128
 Db 61 SFLIIRKLEVDTSIGYICEVEDKRIEVOVLVFLTLAS 98

RESULT 82

AHRB

Ig alpha chain C region - rabbit (fragment)
 C/Species: Oryctolagus cuniculus (domestic rabbit)
 C/Date: 28-Aug-1995 #sequence_revision 28-Aug-1995 #text_change 16-Jul-1999
 C/Accession: A02174
 R/Knight, K.L.; Martens, C.L.; Stoklosa, C.M.; Schneiderman, R.D.
 Nucleic Acids Res. 12, 1657-1670, 1984
 A/Title: Genes encoding alpha-heavy chains of rabbit IgA: characterization of cDNA encod
 A/Reference number: A02174; MUID:84144059; PMID:6322114
 A/Accession: A02174
 A/Molecule type: mRNA
 A/Residues: 1-299 <KNI>
 A/Cross-references: GB:X00353; NID:g1575; PIDN:CAA25100.1; PID:g1576
 C/Comment: This immunoglobulin belongs to the IgA-g subclass. It was isolated from a rab
 C/Complex: An immunoglobulin heterotrimer subunit consists of two identical light (kap
 hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a
 C/Superfamily: immunoglobulin C region, immunoglobulin homology
 C/Keywords: duplication; glycoprotein; heterotrimer; immunoglobulin; plasma
 F:189-261/Domain: immunoglobulin homology <IM2>
 F:138-286/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 8.9%; Score 305; DB 1; Length 299;
 Best Local Similarity 31.6%; Pred. No. 5.2e-11;
 Matches 85; Conservative 44; Mismatches 120; Indels 20; Gaps 11;
 Qy 369 QCLSDSGQVLLBSNIKVLPWTSTPVEPKSCDKHTPCPCPAP--ELLGG--PSVFLFP 423
 Db 20 QCLGKSAACHVEYNSVI--NESLPVPDPCCPANSCTCPSSSRNLSGCGSLSLQIR 77
 Qy 424 PKPDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVNAKTKRREQDYNSTYRVVS 483
 Db 78 PDLGG-LLLGKRDASTLTCTLSGLKNPEDAV-FTW--EPTNGNEPVQORARDLSGCVSVS 133
 Qy 484 VLTVLHQMNLNGKEXKCKVSNKALP-APIETISKAKGQPREPOVYTLPPSRDELTKN-Q 541
 Db 134 VLPSSAEATWKAARTEFTCTVTHPEIDSGSLTATISRGVTP--POVHLPPSEBELALNEQ 191
 Qy 542 VSLTGLVKGFPSPDIAYVESNGQ--PENNY--KTPPVLDSDGSFFLYSKLTVDKSRW 596
 Db 192 VTLTCLVGRGSPKDVLSWMHQGEVDEDSFLVWKMSPESQDKATYAITSLLRVPADW 251
 Qy 597 QCGNVFSCVMHEALHNHYTOKSLSLSPG 625
 Db 252 NQGDYSCVMGHGLAEHFTQRTIDRLAG 280

RESULT 83

S09269

Ig alpha chain C region - rabbit (fragment)
 C/Species: Oryctolagus cuniculus (domestic rabbit)
 C/Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 16-Jul-1999
 C/Accession: S09269
 R/Burnett, R.C.; Hanly, W.C.; Zhai, S.K.; Knight, K.L.
 EMBO J. 8, 4041-4047, 1989
 A/Title: The IgA heavy-chain gene family in rabbit: cloning and sequence analysis of 13
 A/Reference number: S09264; MUID:90076124; PMID:512120
 A/Accession: S09269
 A/Status: not compared with conceptual translation
 A/Molecule type: DNA
 A/Residues: 1-357 <BUR>
 C/Superfamily: immunoglobulin C region, immunoglobulin homology
 C/Keywords: immunoglobulin

F:142-208/Domain: immunoglobulin homology <IM1>

Query Match 8.9%; Score 304.5; DB 2; Length 357;
 Best Local Similarity 33.3%; Pred. No. 7e-11;
 Matches 93; Conservative 39; Mismatches 116; Indels 31; Gaps 15;
 Qy 363 PEAGMOCCLIS---DSGQ-VLBSNIVLPWTSTPVEPKSCDKHTCP-PCPAELLGGP 417
 Db 75 PEENSVAACHVEHNHYDKQHTVPS-----PECCPPTPCPSDTP--TCPCPCPSPS-CGEP 127
 Qy 418 SVLFFPKPDKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVNAKTKRREQDYN 477
 Db 128 SLSLQRFPLLD-LILNSNASLTCTLSGLKNPEDAV-FTW-----EPTNG-NKPVQOSVQS 179
 Qy 478 ----TYRVSVLTVLHQMNLNGKEXKCKVSNKALP-APIETISKAKGQPREPOVYTLPP 532
 Db 180 YPGCGYSVSVLPCCAPPMNAGTEFTCTVTHPEIDGRLTRAKISKDYGAIIPQVHLDP 239
 Qy 533 SRDELTKNQ-VSLTGLVKGFPSPDIAYVESNGQ--QPENNY--KTPPVLDSDGSFFLY 586
 Db 240 PSEELALNELVTLTCLVGRGSPKDVLSWMHQGVNVPENSFVLWKPRLPEPGQEPPTVAVT 299
 Qy 587 SKLTVDKSRWQCGNVFSCVMHEALHNHYTOKSLSLSPG 625
 Db 300 SLLRVPADWNNQNSYTCVVGHEGLAEHFTQRTIDRLSG 338

RESULT 84

S09276

Ig alpha chain C region - rabbit (fragment)
 C/Species: Oryctolagus cuniculus (domestic rabbit)
 C/Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 16-Jul-1999
 C/Accession: S09276
 R/Burnett, R.C.; Hanly, W.C.; Zhai, S.K.; Knight, K.L.
 EMBO J. 8, 4041-4047, 1989
 A/Title: The IgA heavy-chain gene family in rabbit: cloning and sequence analysis of 13
 A/Reference number: S09264; MUID:90076124; PMID:512120
 A/Accession: S09276
 A/Status: not compared with conceptual translation
 A/Molecule type: DNA
 A/Residues: 1-338 <BUR>
 C/Superfamily: immunoglobulin C region, immunoglobulin homology
 C/Keywords: immunoglobulin
 F:228-300/Domain: immunoglobulin homology <IM1>

Query Match 8.8%; Score 302; DB 2; Length 338;
 Best Local Similarity 31.2%; Pred. No. 9.2e-11;
 Matches 84; Conservative 45; Mismatches 120; Indels 20; Gaps 11;
 Qy 369 QCLSDSGQVLLBSNIKVLPWTSTPVEPKSCDKHTPCPCPAP--ELLGG--PSVFLFP 423
 Db 59 QCLGKSAACHVEYNSVI--NESLPVPDPCCPANSCTCPSSSRNLSGCGSLSLQIR 116
 Qy 424 PKPDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVNAKTKRREQDYNSTYRVVS 483
 Db 117 PDLGD-LILGRDASTLTCTLSGLKNPEDAV-FTW--EPTNGNEPVQORARDLSGCVSVS 172
 Qy 484 VLTVLHQMNLNGKEXKCKVSNKALP-APIETISKAKGQPREPOVYTLPPSRDELTKN-Q 541
 Db 173 VLPSSAEATWKAARTEFTCTVTHPEIDSGSLTATISRGVTP--POVHLPPSEBELALNEQ 230
 Qy 542 VSLTGLVKGFPSPDIAYVESNGQ--PENNY--KTPPVLDSDGSFFLYSKLTVDKSRW 596
 Db 231 VTLTCLVGRGSPKDVLSWMHQGEVDEDSFLVWKMSPESQDKATYAITSLLRVPADW 290
 Qy 597 QCGNVFSCVMHEALHNHYTOKSLSLSPG 625
 Db 291 NQGDYSCVMGHGLAEHFTQRTIDRLAG 319

RESULT 85

HVRK2

Ig mu chain C region (clone 12022) - horn shark (fragment)

Ig alpha chain - chicken
C:Species: Gallus gallus (chicken)
C:Date: 18-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 26-Aug-1999
C:Accession: A46507
R:Mansikka, A.
J. Immunol. 149, 855-861, 1992
A>Title: Chicken IgA H chains. Implications concerning the evolution of H chain genes.
A:Reference number: A46507; MUID:92340889; PMID:1634774
A:Accession: A46507
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-585 <MAN>
A:Cross-references: GB:S40610; NID:9251907; PID:9251908
A>Note: sequence inconsistent with the nucleotide translation
A>Note: sequence extracted from NCBI backbone (NCBIN:109906, NCBIPI:109907)
C:Superfamily: immunoglobulin C region; immunoglobulin homology
F:34-117//Domain: immunoglobulin homology <IMM>

Query Match 8.7%; Score 295.5; DB 2; Length 585;
Best Local Similarity 24.3%; Pred. No. 4.5e-10;
Matches 161; Conservative 87; Mismatches 264; Indels 151; Gaps 35;

```
QY 10 LLLVQLALPRAAOGNVLGKK--GDTVELTCTAS--QKKSIOFHH-----KNSNQI 59
DB 8 LLLLAALGLMAAVTLTDSGGGLQTPGALSLVCKASFTTSYSWMGMRAQPGGLEWV 67
QY 60 KILNQSGFLTRKPSKLNDRADSRSLWDQGNFPLIKNLKIEDSDTYICEVEDQKEEVQ 119
DB 68 AGIDDDSGTGCGVAVQGRATISRDN--GQSTVRLQLNMLAEDPATYCYCKA----- 118
QY 120 LLVRLGTLNSDTHLIQGGSLTTLTLESPPGSSPSV-----QCRSPRKNIQGGKTLV 171
DB 119 --ASGCGACASIDAMGHTGTEIVTSSASASRPTLQLPLPSDCDP---VNTIG-CLVT 172
QY 172 SOLELDGSGTCTYLOQKQKVEFKIDIVLAFOKASSIYKKBEQVBFSPPLATFYEK 231
DB 173 SFLPPVTVTWTTGGAAATLV-----TSLPATGTGIVSLTTLALTVPEEQ 218
QY 232 LTGSGELMMQAEARASSSKSWITFDLKNK-----EVSVKRVT--QDPKLOMGKLLPLHL 282
DB 219 LQGN-EFVCRQAHATGA-----DYKERTIGDGVCPIRTSKVTLTSDTQDFFERVLV 272
QY 283 TLPOALPOYAGSGNLTALAEKTKLHOENVLVNRAATOLQKNLTCEVWGPTSPKLMSL 342
DB 273 CLVEGLPS-AGA-----AIQ-----W----- 287
QY 343 KLEKAEKAVSKREKPVWVLNPEAGMOCCLSDSGOVLESNIKVL-PTWSTPVEPKSCDK 401
DB 288 -LQDNEKTPAPE-----SDSGGSDC--TESG-VTQMSRVNVTYRKSMEGAAQ-FGCRV 336
QY 402 TH-----TCPPCPA-PELLGGPSVFLPPKPKDTLMISRTPEVTCVAVVDVSHEDP 450
DB 337 THGALKEPVTAIVTSDCATPOL-----QVSLPLPLEE-LVSHNATYTCVSNAAAAD- 390
QY 451 EVKENWYVD---GVEVNAKTKPREEOYNSTYRVSVLTVLHQMNLNGEKYCKVSNKAL 507
DB 391 GVSWSMSRSSGGGLDV-----SQTEDRQADGRYTVASFRLVCAEENNGETGCGSVREGV 446
QY 508 PAPLEKTSKAKGP-REPOVYTLPPSRDELTL-KNOVSLTCLVKGFPYSDIIVAVMESNGQ 565
DB 447 -VVAESIRKETDTPPLHAPSVYFPPPAEELSLQETATLTMASSFSSILLTWTOQNQ 505
QY 566 P-ENNYKTPPVLDSDGSFF-LYSKLTVDKSRMQQGNVFCSVMEALNHNHYTKSISL 622
DB 506 PISQNYILIFGP--EKQDGFYSLYKLVSVEDMQRGVFGCVGHDOIPLNFIHKSIDK 563
QY 623 SPG 625
DB 564 NAG 566
```

RESULT 89
HVRCO

Ig mu chain C region, membrane-bound (clone 3050) - horn shark
C:Species: Heterodontus francisci (horn shark)
C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Aug-1996
C:Accession: S01854; C32716; A66530
R:Kokubu, F.; Hinds, K.; Litman, R.; Shambloct, M.J.; Litman, G.W.
EMBO J. 7, 1979-1988, 1988
A>Title: Complete structure and organization of immunoglobulin heavy chain constant region
A:Reference number: S00980; MUID:88328985; PMID:3138109
A:Accession: S01854
A:Molecule type: DNA
A:Residues: 1-461 <KOK>
A:Cross-references: EMBL:X07781
A>Note: the sequence was determined from the germ-line gene
R:Kokubu, F.; Hinds, K.; Litman, R.; Shambloct, M.J.; Litman, G.W.
Proc. Natl. Acad. Sci. U.S.A. 84, 5868-5872, 1987
A>Title: Extensive families of constant region genes in a phylogenetically primitive vert
A:Reference number: A32716; MUID:87289703; PMID:3475706
A:Accession: C32716
A:Molecule type: DNA
A:Residues: 1-99 <K02>
A:Cross-references: GB:M17186
C:Genetics:

A:Introns: 100/1, 206/1, 309/1, 419/1, 459/3
C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobulin; immunoglobulin homology <IMM1>
F:123-190/Domain: immunoglobulin homology <IMM2>
F:228-231/Domain: immunoglobulin homology <IMM3>
F:330-400/Domain: immunoglobulin homology <IMM4>
F:438-458/Domain: transmembrane #status predicted <TM>
F:27-85,130-188,235-289,337-398/Disulfide bonds: #status predicted
F:164,200,245,374,411,415,437/Binding site: carbohydrate (Aan) (covalent) #status pr

Query Match 8.6%; Score 295; DB 1; Length 461;
Best Local Similarity 25.5%; Pred. No. 3.6e-10;
Matches 121; Conservative 69; Mismatches 167; Indels 118; Gaps 24;

```
QY 208 SSIYKKEGQVER---SFLPAF-----VEKLTGSGELMMQAEARASSKSW 251
DB 37 TSVSMKQNEBITTGLKTPSVLNKKGTYYQSSQLTTSEVSSKTYCERREGS--VW 94
QY 252 ITFDLKNKEVSVKRVNTPDPLQMGKLLPLHLTPQALPOYAGSGNLTALAE----- 303
DB 95 I-----KEI-----PDCK---GDKVHPVVIILQSSSEITSRFATVLCISIDFPES 139
QY 304 -----KTGLHQENVLVNRAATOLQKNLTCEVWGPTSPKLMSLKLEKAEKVKREKP 357
DB 140 ITVSMWKDQO-HMESGFV-----SPTCGVNGTFSATSRLT-----VPARE-- 179
QY 358 VWVLNPEAGMOCCLSDSGOVLESNIKVLPTWSTPVEPKSCDKTHCPPCPAPELGGP 417
DB 180 -WFTNK--VYTCQVSHQG-VTQSRNI-----TSGQVPSQCDN---P 213
QY 418 SVFLPPKPKDTLMISRTPEVTCVAVVDVSHEDVEKENWYVDGVEVNAKTKPREEOYNS 477
DB 214 VIKLPPSIEQVL-LEATVTLTCV---VSNAPGVNVMS--TQEQKSLSEIIVAGGED 266
QY 478 TYRVSVLTVLHQMNLNGEKYCKVSNKALPAPLEKTSKAKGP-REPOV-YTLPPSRD 535
DB 267 ADVSISTVNISTQAMLSGAEFYCVVNHQDLPPLRASIHKEEVKDLREPSVILLSPAED 326
QY 536 ELTNGVSLTCIVKGFPSDIAVEMESNGQPEN--NYKTPPVLDSD-GSPFLYSKLTVD 592
DB 327 VSAQRFLSLTCLQGFPPREIFVKMTVNDKSVNPGNKNTBVMANENNSFTYLSLSIA 386
QY 593 KSRWQGNVFCSVMEALNHNHYTKSISLPGIQLDETCAEADGDLGIMTDD 647
DB 387 ABEWASGASVSCVGHAI-----PLKIINRTYVKKSSDS-DHIMIED 428
```

RESULT 90

Db 303 ESVCVVAHEGLASHFTQRTIDRLAG 328

RESULT 93

WCHH

Ig mu chain C region - chicken (fragment)

C/Species: Gallus gallus (chicken)

C/Date: 03-Aug-1984 #sequence_revision 03-Aug-1984 #text_change 22-Jun-1999

C/Accession: A02170

R:Dahan, A.; Reynaud, C.A.; Weill, J.C.

Nucleic Acids Res. 11, 5381-5389, 1983

A/Title: Nucleotide sequence of the constant region of a chicken mu heavy chain immunoglobulin

A/Reference number: A02170; MUID:83299221; PMID:6310496

A/Accession: A02170

A/Molecule type: mRNA

A/Residues: 1-367 <DAH>

A/Cross-references: GB:K00389; NID:g212204; PID:AAA48923.1; PID:g212205

C/Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a

C/Superfamily: immunoglobulin C region; immunoglobulin homology

C/Keywords: duplication; glycoprotein; heterotetramer; immunoglobulin

F/49-133/Domain: immunoglobulin homology <IMM2>

F/151-219/Domain: immunoglobulin homology <IMM3>

F/158-329/Domain: immunoglobulin homology <IMM4>

F/347-367/Domain: carboxyl-terminal <CTS>

F/51.119.303.354/binding site: carbohydrate (Asn) (covalent) #status predicted

F/156-111.158-217.265-327/disulfide bonds: #status predicted

F/126,366/disulfide bonds: interchain (to mu chain) #status predicted

F/205/disulfide bonds: interchain (to mu chain in another subunit) #status predicted

Query Match 8.5%; Score 291; DB 1; Length 367;

Best Local Similarity 30.8%; Pred. No. 4.6e-10;

Matches 91; Conservative 48; Mismatches 126; Indels 30; Gaps 14;

Qy 354 REKP--VWVUN--PEAGMOCCLSDSGVLESNIKLP--WSTPV-----EPKSC 399

Db 61 RRRPLEVWVWYKNGSVAAATATATVGEVVAESRISVTESEMPGTATFSCVVEGEMNT 120

Qy 400 DKTHTC--PCGPAPELGSGSVFLFPPPKDTLMISRPETCVVVDVSHEDP--EVKPNW 456

Db 121 SKRMECGLEPVQOYI---AIRYTPSFVD-IFISKSATLTCSRYSNNVADGLEVSW-W 174

Qy 457 YVDGEVNAKTKPREEOYNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIEKTIS 516

Db 175 KEKGKULETALGK-RVQSGNGLYTDGVAITCASMDGDDGYCKVNPDLLEFMEEKMR 233

Qy 517 KAK-GQPREPOVYTLPSRDELTKNQ-VSLTCLVKGFYPSDIAVEWESNGQ--PENNYKT 572

Db 234 KTKASNAAPPSVYVFPPTTEOLINGNQRLSVTCMAQGFNPRLFVRMMKNGEPLFQSGSVT 293

Qy 573 TRPVLD--DGSFLYSKLTVDKSRWQGNVFSQSVMEHALHNYTOKSLSLSPG 625

Db 294 SAPMENENESVAVSVLVGAEEMGAGNVYTLVGHLEALPLQLAKSVDRASG 348

RESULT 94

S09272

Ig alpha chain C region - rabbit (fragment)

C/Species: Oryctolagus cuniculus (domestic rabbit)

C/Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 16-Jul-1999

C/Accession: S09272

R:Bunnett, R.C.; Hanly, W.C.; Zhai, S.K.; Knight, K.L.

EMBO J. 8, 4041-4047, 1989

A/Title: The IgA heavy-chain gene family in rabbit: cloning and sequence analysis of 13

A/Reference number: S09264; MUID:90076124; PMID:2512120

A/Accession: S09272

A/Status: not compared with conceptual translation

A/Molecule type: DNA

C/Superfamily: immunoglobulin C region; immunoglobulin homology

C/Keywords: immunoglobulin

F/232-305/Domain: immunoglobulin homology <IMM>

Query Match 8.4%; Score 285.5; DB 2; Length 343;

Best Local Similarity 33.9%; Pred. No. 8.9e-10;

Matches 81; Conservative 35; Mismatches 100; Indels 23; Gaps 12;

Qy 399 CDKHTT-CPPCAPDELGFSVFLFPPPKDTLMISRPETCVVVDVSHEDP--VKPNW 456

Db 97 CQDCHCYCPPTSC---GEPSSLQRPDIDLLLESKA-SLTCTLSGL--KDPGAVFTW 149

Qy 457 Y-VDGEVNAKTKPREEOYN-STYRVSVLTVLHODMLNGEKYCKVSNKALP-APIEK 513

Db 150 EPTNG---NEFVQGSVQSTPPCGYSVSVLPGCAEFMNAETCTVTHPEIGSSLTA 205

Qy 514 TISKAKQPREPOVYTLPSRDELTKNQ-QVSLTCLVKGFYPSDIAVEWESNGQ--PENN 569

Db 206 TISRSRSLTPRPVHLLPPTTEALNEQVTLTCLVGFSPKQVLYSWTHNGTLVPEKDS 265

Qy 570 Y---KTPPVLDSDGSFLLYSKLTVDKSRWQGNVFSQSVMEHALHNYTOKSLSLSPG 625

Db 266 YLWVKPLPEPGQPTTYAVTSLRVSAMEDWNGDSYSCVVGHGGLAHEFTQRTIDRQAG 324

RESULT 95

S12328

Ig heavy chain C region (clone 5301) - horn shark (fragment)

C/Species: Heterodontus francisci (horn shark)

C/Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 21-Jan-2000

C/Accession: S12328; S01855

R:Litman, R.

Submitted to the EMBL Data Library, May 1988

A/Reference number: S12328

A/Accession: S12328

A/Molecule type: mRNA

A/Residues: 1-244 <LIT>

A/Cross-references: EMBL:X07785; NID:g633965; PID:CAA0618.1; PID:g633966

A/Note: This sequence was determined from the differentiated gene

R:Kokubu, F.; Hinds, K.; Litman, R.; Shambloct, M.O.; Litman, G.W.

EMBO J. 7, 1979-1988, 1988

A/Title: Complete structure and organization of immunoglobulin heavy chain constant reg1

A/Reference number: S00980; MUID:88328985; PMID:3138109

A/Accession: S01855

A/Molecule type: mRNA

A/Residues: 200-244 <KOK>

A/Cross-references: EMBL:X07785

A/Note: This sequence was determined from the differentiated gene

C/Superfamily: immunoglobulin C region; immunoglobulin homology

C/Keywords: glycoprotein; heterotetramer; immunoglobulin; transmembrane protein

F/11-244/Domain: C region (fragment) <CRE>

F/11-181/Domain: immunoglobulin homology <IMM>

F/215-241/Domain: transmembrane #status predicted <TM>

F/26,155,192,196/binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 8.3%; Score 284; DB 2; Length 244;

Best Local Similarity 32.2%; Pred. No. 7e-10;

Matches 75; Conservative 39; Mismatches 89; Indels 30; Gaps 9;

Qy 428 DTLMISRPETCVVVDVSHEDPEVKPNWYVDGEVNAKTKPREOY-----NSTRVV 482

Db 4 EQVILEATVLTTCV--VSNAPYGVANVSW-----TQCKPLKSIANVQPGDSISVI 52

Qy 483 SVLTVLHODMLNGEKYCKVSNKALPAPIEKTISKAGQP-REPOV-YTLPPSRDELTKN 540

Db 53 STYDISAQSWLSGDVFCVSVSHDLPFLRDFIHKEKNKDLREBSVVLPPADVAGQR 112

Qy 541 QVSLTCLVKGFYPSDIAVEWESNGQPN--NYKTPPVLDSD-GSFLLYSKLTVDKSRWQ 597

Db 113 FLSTLTCLVRGFSRRELTVKTVNDKSVNPNGNKYNTETVMAANDSSYFISLSIADEWA 172

Qy 598 QGNVFSQSVMEHALH-----NHYTOKSLSLSPGLDTECAEOGDELGLWTT 646

Db 173 SGASYSQVGHLEALPKTIKRYVKKSSDS-----DHVIEDNEEESGNITWT 220

```

RESULT 96
B22360
Ig alpha-2 chain C region (allotype A2m(1)) - human
C:Species: Homo sapiens (man)
C>Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 16-Jul-1999
C:Accession: B22360
R:Flanagan, J.G.; Lefranc, M.P.; Rabbitts, T.H.
Cell 36, 681-688, 1984
A:Title: Mechanisms of divergence and convergence of the human immunoglobulin alpha-1 ar
A:Reference number: A94653; MUID:84130179; PMID:6421489
A:Accession: B22360
A:Molecule type: DNA
A:Residues: 1-340 <FLA>
C:Gene: GDB:IGHA2
A:Cross-references: GDB:119333; OMIM:147000
A:Map position: 14q32.33-14q32.33
A:Introns: 1/1 103/1; 210/1
C:Superfamily: Immunoglobulin C region; Immunoglobulin homology
C:Keywords: Immunoglobulin
F:230-302/Domain: Immunoglobulin homology <IMM>

Query Match      8.3%; Score 282.5; DB 2; Length 340;
Best Local Similarity 28.3%; Pred. No. 1.3e-09;
Matches 100; Conservative 47; Mismatches 137; Indels 69; Gaps 17;

QY 333 PTPSKLM-LSLKLENKAKVSKR-----EKPVVVLNPEAGM-----WQCCLSDSQ 377
DB 3 PTPSPKFLSLDSTQDGNVAVVACLVQGFPOEPISVTSESQNTANFPPSDASGD 62
QY 378 VLLSNIKVLPWTSPVEPKS--CD-KYHTCP-----PCPAPELLGSPVLFPPKP-- 426
DB 63 LYTTSSQLTLPATNCGP-DEKSVTCVHKYHTNSQDVTVPCFVP-----PPPPCC 110
QY 427 -----KDTLMSRPEVTCVYVVDVSHEDPEVKFNWYDGVENAKTKPPEQYN 476
DB 111 HPRLLHRPALDILLGSEBANLTCTLTGL-RDASGATFTWTPSSGK--SAVGGPPEBDLC 167
QY 477 STYRVSVLTVLHODMLNGKEYKCKVSNKALPAPIEKTISSKAKGQPREPOVYTLPPSDE 536
DB 168 GCIYSVSVLPSCAQPMNHETFTCTAAHPELKTPLTANITKS-GNTFREYVHLPPSDE 226
QY 537 LTKNO-VSLTCLVKGFPSDIAVENESNQ--PENNYKTPPEVLD-SDG--SFPLYSKLT 590
DB 227 LALNELVLTCLARGPSKPDVLRMLQSGQLPREKYLTMASROPSQTTFAATSLIR 286
QY 551 VDKSMQCGNVSCSVMEALHNHYTKSLSPG-----LQIDETC 632
DB 287 VAABDMKKGDTFSCWGHLEPLAFTQKTRIDLAGKPTHVNVSVMAEVDGTC 339

RESULT 97
HYRKC1
Ig mu chain C region (clone 6121) - horn shark (fragment)
C:Species: Heterodontus francisci (horn shark)
C>Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jul-1999
C:Accession: S01852
R:Kokubu, F.; Hinde, K.; Litman, R.; Shambloot, M.J.; Litman, G.W.
EMBO J. 7, 1979-1988, 1988
A:Title: Complete structure and organization of immunoglobulin heavy chain constant regi
A:Reference number: S00980; MUID:88328985; PMID:3138109
A:Accession: S01852
A:Molecule type: mRNA
A:Residues: 1-393 <KOK>
A:Cross-references: EMBL:X07782; NID:963961; PIDN:CAA30615.1; PID:9833623
A:Note: The sequence was determined from the differentially expressed gene
C:Complex: An immunoglobulin heterodimer subunit consists of two identical light (kap
chain) disulfide bonds. In some cases, such as IGA and IGM, the subunits associate into 1a
C:Superfamily: Immunoglobulin C region; Immunoglobulin homology
C:Keywords: duplication; glycoprotein; heterodimer; immunoglobulin
F:1-393/Domain: C region (fragment) <CRB>
F:1-42/Domain: Immunoglobulin homology (fragment) <IM1>
F:78-145/Domain: Immunoglobulin homology <IM2>

```

```

F:183-246/Domain: Immunoglobulin homology <IM3>
F:285-355/Domain: Immunoglobulin homology <IM4>
F:119,155,200,230,329,366,370,380/Binding site: carbohydrate (Asn) (covalent) #status pr

Query Match      8.2%; Score 281.5; DB 1; Length 393;
Best Local Similarity 26.7%; Pred. No. 1.8e-09;
Matches 116; Conservative 63; Mismatches 164; Indels 91; Gaps 22;

QY 211 VYKKEGEVSEFPFLAFVTEKLTGSGELMQAERASSKSWTFPLKNKEVSKRVTD 270
DB 13 VLNKGTYTQ--SQQLTIT-ESEVSSKTYCEVRGES--VLT-----KEI-----PDC 56
QY 271 KLQMGKULPLHLTLPQALPOYAGSGNLTALAE-----XTGKLHGVNLV 316
DB 57 K--GDKNHPYVILVQSSSEITSRFAVTLCSIIDHPESITVSWLMDQ--HMESGFT 112
QY 317 MRATQLOKLVCEWGPSPKLMLSLKENKAKVSKREKPVVVLNPEAGMQLSDSG 376
DB 113 -----SPICGVNGTFSATSRLT-----VPAE--WFPTNK--VYTCQVSHQ 149
QY 377 QVLESNIKVLPTWSTPVEPKSCDKTHTCPCPAPELLGSPVLFPPKPDMLMISRT 436
DB 150 -VTQSRNL-----TGSQVPCND-----PVIKLPPSIEQVL-LEATV 186
QY 437 EYTCVAVVDVSHEDPEVKFNWYDGVENAKTKPREEOYNSTYRVSVLTVLHODMLNGK 496
DB 187 TLNCL--VSNAPYVNVSW---TQEQSLKSELAIVQGEDADSVISTVNIStOMLSGA 240
QY 497 EYCKVSNKALPAPIEKTISSAKGP-REPOV-YTLPPSRDELTKNOVSLTCLVKGFPS 554
DB 241 EYFCVAVNHQDPTPLRASIHKEEVKDLREPVSILISPAEDVSAQRFLSTCLVGVFSPR 300
QY 555 DIAVWENSGQEN--NKTTPEVLDSD-GSFPLYSKLTVDKSRQCGNVSCSVMEAL 611
DB 301 EIVKWTINDSVMNGNKTEVMAENDSSVFYLSLISAAEWAASGASVSCVGHBAI 360
QY 612 HNHVYOKSLSPG 625
DB 361 PLKIKRTVKNSSG 374

RESULT 98
S21462
T-cell surface glycoprotein CD4 (allele 2) - pig (fragment)
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 20-Feb-1995 #sequence_revision 19-Apr-1996 #text_change 21-Jan-2000
C:Accession: I47132; S21462
R:Gustafsson, K.; Germana, S.; Sundt, T.M.
J. Immunol. 151, 1365-1370, 1993
A:Title: Extensive allelic polymorphism in the CDR2-like region of the miniature swine C
A:Reference number: I47131; MUID:93329116; PMID:8335933
A:Accession: I47132
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-99 <GU2>
A:Cross-references: EMBL:X65630; NID:91929; PIDN:CAA46584.1; PID:9388233
C:Superfamily: T-cell surface glycoprotein CD4; Immunoglobulin homology
C:Keywords: glycoprotein; T-cell
F:3-81/Domain: Immunoglobulin homology <IMM>

Query Match      8.2%; Score 280.5; DB 2; Length 99;
Best Local Similarity 56.1%; Pred. No. 3.6e-10;
Matches 55; Conservative 19; Mismatches 23; Indels 1; Gaps 1;

QY 32 KKGDTVELTCTASQKSIQFMKNSNQIKILNGSGFLTKGP-SKLNDRADRSRLMDQ 90
DB 1 KAGDLAEPLCHSSQKQKLPFWKNSDQKILSRNLMHKASVTELSRLDSKKQMDHG 60
QY 91 NFPLIKLKLEDSQTYICEVEDQKEVQLLVFGLTAN 128
DB 61 SFPLIKKLEVTDSGIYICEVEDKRIEVOVLVFLTLAS 98

```

```

RESULT 99
Ig alpha chain C region - rabbit (fragment)
C/Species: Oryctolagus cuniculus (domestic rabbit)
C/Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 16-Jul-1999
C/Accession: S09266
R/Burnett, R.C.; Hanly, W.C.; Zhai, S.K.; Knight, K.L.
EMBO J. 8, 4041-4047, 1989
A/Title: The IGA heavy-chain gene family in rabbit: cloning and sequence analysis of 13
A/Reference number: S09264; MUID:90076124; PMID:2512120
A/Accession: S09266
A/Status: not compared with conceptual translation
A/Molecule type: DNA
A/Residues: 1-352 <BUR>
C/Superfamily: immunoglobulin C region; immunoglobulin homology
C/Keywords: immunoglobulin
F/241-314/Domain: immunoglobulin homology <IMM>

Query Match      8.2%; Score 280.5; DB 2; Length 352;
Best Local Similarity 32.2%; Pred. No. 1.8e-09;
Matches 85; Conservative 37; Mismatches 109; Indels 33; Gaps 13;

Qy 374 DSGVLLSNIKVLPTWSTPEKSCDKHTCPCPAPDEL-LGDPVSFLPFPKPKDTLMI 432
Db 91 DEGNL-----TVLYPECKDPSDPTCPCPPTCGEPLSLQRPDID-LTL 137
Qy 433 SRTEVTCVVVDVSHEDPE-VKFNMY-VDGVEVFNNAKTKREEQDYN-STYVVSFLTYM 489
Db 138 ESNASLTTLTSLG-L-KDEGAVFTWNPNTNGNEFVQOST---QSPCCGCVSVSLPQCA 191
Qy 490 QDMNGKEKCKVSNKALP-APIETKISKAGQREPOVYTLPSRDELTKNQ-VSLTCL 547
Db 192 EPMNAGTEFTCTVHPELGGSLRTATISRSGLTP-PQVHLPLPPTTELANALVTILCL 249
Qy 548 VKGPEPSDIAVESNSGQ--PENNY--KTPPYLSDSGSFYLSKLTVDKSRMQGNV 601
Db 250 VRGSPKVLVSWTHNGTLVVPKOSFLWKPLPEPGCEPTVAVTSLRLVAEDMNQDGS 309
Qy 602 FSGVMHEALNNHTOKSLSPG 625
Db 310 YSCVGHGHEGLAEHFTQRTIDRLAG 333

RESULT 100
A/HU
Ig alpha-1 chain C region - human
C/Species: Homo sapiens (man)
C/Date: 22-May-1991 #sequence_revision 03-Oct-1995 #text_change 20-Oct-2000
C/Accession: A22360; A92249; A91662; S38979; B53110; A02171
R/Flanagan, J.G.; Lefranc, M.P.; Rabbitts, T.H.
Cell 36, 681-688, 1984
A/Title: Mechanisms of divergence and convergence of the human immunoglobulin alpha-1 ar
A/Reference number: A94653; MUID:84130179; PMID:6421489
A/Accession: A22360
A/Molecule type: DNA
A/Residues: 1-353 <FLA>
R/Putnam, F.W.; Liu, Y.S.V.; Low, T.L.K.
J. Biol. Chem. 254, 2865-2874, 1979
A/Title: Primary structure of a human IgA1 immunoglobulin. IV. Streptococcal IgA protea
A/Reference number: A92249; MUID:79151016; PMID:107164
A/Contents: myeloma protein Bur; disulfide bonds
A/Accession: A92249
A/Molecule type: protein
A/Residues: 1-16, 'Z', 18, 'B', 20, 'B', 22-34, 'Q', 36-45, 'Z', 47-51, 'B', 53-56, 'ZB', 59-61, 'B', 63
303, 'B', 305-346, 'Q', 348-353 <PUT>
A/Note: this is the final paper in a series
R/Kratz, H.; Alevoigt, P.; Ruban, E.; Starescik, K.; Hilschmann, N.
Hoppe-Seyler's Z. Physiol. Chem. 356, 1337-1342, 1975
A/Title: The primary structure of a monoclonal IgA-immunoglobulin (IgA Tro.), II: the ar
A/Reference number: A91662; MUID:76023781; PMID:809331
A/Accession: A91662
A/Molecule type: protein

```

```

A/Residues: 1-16, 'Z', 18, 'B', 20, 'B', 22-34, 'Q', 36-45, 'Z', 47-51, 'B', 53-56, 'ZB', 59-61, 'B', 63-
, 'R', 232-237, 'Q', 240-243, 'Q', 245-283, 'Q', 285-289, 'E', 291-303, 'B', 305-353 <KRA>
A/Experimental source: myeloma protein Tro
R/Falgreen-Gebauer, E.; Gebauer, W.; Baetian, A.; Kratzin, H.D.; Eilfert, H.; Zimmermann
Biol. Chem. Hoppe-Seyler 374, 1023-1028, 1993
A/Title: The covalent linkage of secretory component to IgA. Structure of sIgA.
A/Reference number: S38978; MUID:94121784; PMID:8292260
A/Accession: S38978
A/Molecule type: protein
A/Residues: 188-196, 'D', 198-201 <PAL>
R/Iang, C.T.; Kratzin, H.; Goetz, H.; Hilschmann, N.
Hoppe-Seyler's Z. Physiol. Chem. 360, 1919-1940, 1979
A/Title: Die Primärstruktur eines monoklonalen IgA1-Immunglobulins (Wydolproteins Tro).
A/Reference number: A91684; MUID:80114124; PMID:393607
A/Contents: annotation; Tro; disulfide bonds
A/Note: Cys-14 bonds to a light chain
R/Calejo, M.; Bescibano, U.; Grubb, A.; Mendez, E.
J. Biol. Chem. 269, 384-389, 1994
A/Title: Location of a novel type of interpeptide chain linkage in the human protein
A/Reference number: A53110; MUID:94103241; PMID:7506257
A/Accession: B53110
A/Molecule type: protein
A/Residues: 346-351, 'X', 353 <CAL>
C/Genetics:
A/Gene: GDB:IGHA1
A/Map position: 14q32.33-14q32.33
A/Intons: 1/1; 103/1; 223/1
C/Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa)
chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la;
C/Superfamily: immunoglobulin C region; immunoglobulin homology
C/Keywords: chromoprotein; duplication; glycoprotein; heterotetramer; immunoglobulin; tr
F/140-206/Domain: immunoglobulin homology <IM1>
F/243-315/Domain: immunoglobulin homology <IM2>
F/26-85, 77-101, 123-180, 147-204, 250-313/Disulfide bonds: #status experimental
F/105, 111, 113, 119, 121/Binding site: carbohydrate (Ser) (covalent) #status experimental
F/122, 182/Disulfide bonds: interchain #status experimental
F/144, 340/Binding site: carbohydrate (Asn) (covalent) #status experimental
F/192/Disulfide bonds: interchain (to secretory component) (partial) #status experimental
F/192/Binding site: cysteine (Cys) (covalent) (partial) #status experimental
F/192/Disulfide bonds: interchain (partial) #status experimental
F/352/Cross-link: alpha-1-microglobulin-Ig alpha complex chromophore (Cys) (interchain to
F/352/Disulfide bonds: interchain (to J chain) (partial) #status experimental
F/352/Disulfide bonds: interchain (partial) #status experimental

Query Match      8.2%; Score 280.5; DB 1; Length 353;
Best Local Similarity 26.8%; Pred. No. 1.8e-09;
Matches 93; Conservative 49; Mismatches 132; Indels 73; Gaps 13;

Qy 305 TGKLGQVNLVWRATQ--LQKNLTCEVWGPTSPKMLSLKENKAKVSKREKPVVVLN 362
Db 60 SGLDLYTSSQLTPATDCLAGKSVTCVHKYTPSQVYTP----- 100
Qy 363 PEAGMOCCLSDSGVLLSNIKVLPTWSTPEKSCDKHTCPCPAPDELGGPSVFLF 422
Db 101 -----CPVSTPPTP-----SPSTPPTPSGCC-HPLSLH 130
Qy 423 PPKPKDTLMSRPEVTCVVVDVSHEDPEVKFNMYVDGVEVFNNAKTKREEQNSTRVV 482
Db 131 RPALED-LILGSBANLITCTLGL-RDASGTFWTPESSGR--SAVQGPERRDLGCGSVS 186
Qy 483 SVLTVHODPLNGEKYCKVSNKALPAPIETKISKAGQREPOVYTLPSRDELTKNQ- 541
Db 187 SVLPGCCPEPNHGTFTCTVAYPEKSTPLPATLSKS-GNTFRREVHLPPSEELANEL 245
Qy 542 VSLTCLVKGPEPSDIAVESNSGQ--PENNYKTPPYLSD--SG--SFYLSKLTVDKSRW 596
Db 246 VTILCLARGSPDVAVRWLQSGQELPREKYLTMASQGESQGTTFVAVSILVAEDW 305
Qy 597 QQGVNFSQSVMEALNNHTOKSLSPG-----LQUDETC 632
Db 306 KKGDTECMVGHGHEALPLAFQKTIIDRLAGKPTVNVSVVAEVDGTC 352

```

Search completed: August 3, 2004, 13:14:49
Job time : 20.8924 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 12:50:44 ; Search time 8.8991 Seconds
(without alignments)
3791.557 Million cell updates/sec

Title: SEQ3
Perfect score: 3414
Sequence: 1 MNRGVPRHLLVLQLALLP.....DETCAADQDELGLMTTDP 648

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 125 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2024	59.3	458	1 CD4_HUMAN	P01330 homo sapien
2	2000	58.6	458	1 CD4_PANTR	P16004 pan troglod
3	1853	54.3	458	1 CD4_MACFU	P79184 macaca fusc
4	1852	54.2	458	1 CD4_MACMU	P16003 macaca mul
5	1849	54.2	458	1 CD4_MACFA	P79185 macaca fasc
6	1844	54.0	458	1 CD4_MACNE	Q08340 macaca neme
7	1827	53.5	458	1 CD4_CERAE	Q08338 cercopithe
8	1726	50.6	397	1 CD4_CERPT	Q08336 cercocobu
9	1716	50.3	397	1 CD4_CERTO	Q08337 saimiri sci
10	1581.5	46.3	457	1 CD4_SATSC	P01857 homo sapien
11	1266	37.1	330	1 GC1_HUMAN	P01860 homo sapien
12	1151	33.7	290	1 GC1_HUMAN	P01859 homo sapien
13	1150	33.7	326	1 GC2_HUMAN	P46630 oryctolagus
14	1149	33.7	459	1 CD4_RABIT	P01861 homo sapien
15	1141	33.4	327	1 GC4_HUMAN	P33705 canis fami
16	1141	33.4	463	1 CD4_CANFA	P05540 rattus norv
17	999	29.3	457	1 CD4_RAT	P05632 mus musculi
18	993	29.1	457	1 CD4_MOUSE	P03887 mus musculi
19	947.5	27.8	398	1 GC3_MOUSE	P01869 mus musculi
20	940.5	27.5	393	1 GC1_MOUSE	P01870 oryctolagus
21	922	27.0	323	1 GC_RABIT	P01865 mus musculi
22	915.5	26.8	399	1 GC4_MOUSE	P01862 mus musculi
23	906.5	26.6	329	1 GC2_MOUSE	P01867 mus musculi
24	884.5	25.9	405	1 GC3_MOUSE	P22436 mus musculi
25	849.5	24.9	329	1 GC4_MOUSE	P20761 rattus norv
26	844	24.7	333	1 GC1_MOUSE	P20759 rattus norv
27	828	24.3	326	1 GC1_MOUSE	P01868 mus musculi
28	826.5	24.3	326	1 GC1_MOUSE	P20762 rattus norv
29	817.5	23.9	329	1 GC2_MOUSE	P01864 mus musculi
30	816	23.9	330	1 GC3_MOUSE	P01863 mus musculi
31	814.5	23.9	335	1 GCA_MOUSE	P20760 rattus norv
32	795	23.3	322	1 GCA_MOUSE	P01866 mus musculi
33	788.5	23.1	336	1 GCB_MOUSE	

34	375.5	11.0	476	1 MUCM_MOUSE	P01873 mus musculi
35	364.5	10.7	421	1 EPC_MOUSE	P06336 mus musculi
36	364	10.7	455	1 MUC_MOUSE	P01872 mus musculi
37	359.5	10.5	454	1 MUC_HUMAN	P01871 homo sapien
38	359	10.5	428	1 EPC_HUMAN	P01854 homo sapien
39	358.5	10.5	479	1 MUCM_RABIT	P04221 oryctolagus
40	355.5	10.4	429	1 EPC_RAT	P01855 rattus norv
41	355.5	10.4	454	1 MUC_MESAU	P06337 mesocricetu
42	353.5	10.4	391	1 MUCB_HUMAN	P04220 homo sapien
43	349	10.2	458	1 MUC_FABIT	P03988 oryctolagus
44	348	10.2	457	1 MUC_SUNMU	P20768 suncus murt
45	340	10.0	450	1 MUC_CANFA	P01874 canis fami
46	305	8.9	299	1 ALC_RABIT	P01879 oryctolagus
47	298	8.7	446	1 MUC_CHICK	P01875 gallus gall
48	297	8.7	438	1 HVC2_HETPR	P23085 heterodontu
49	295	8.6	461	1 HVC1_HETPR	P23088 heterodontu
50	293.5	8.6	438	1 HVC3_HETPR	P23087 heterodontu
51	282.5	8.3	340	1 ALC2_HUMAN	P01877 homo sapien
52	281.5	8.2	393	1 HVC3_HETPR	P23086 heterodontu
53	280.5	8.2	353	1 ALC1_HUMAN	P01876 homo sapien
54	278	8.1	353	1 ALC1_GORGO	P20758 gorilla gor
55	267	7.8	370	1 HVC1_HETPR	P23084 heterodontu
56	265.5	7.8	481	1 MUCM_ICTPU	P23735 ictalurus p
57	248.5	7.3	344	1 ALC_MOUSE	P01878 mus musculi
58	203.5	6.0	513	1 SHS1_MOUSE	P97797 m protei
59	187.5	5.5	506	1 SHS1_BOVIN	O46631 bos taurus
60	182.5	5.3	739	1 VCA1_RAT	P22534 rattus norv
61	178	5.2	4391	1 PGBM_HUMAN	P98160 homo sapien
62	177.5	5.2	105	1 LAC1_MOUSE	P01843 mus musculi
63	174.5	5.1	6632	1 UN89_CAEBL	O01761 caenorhabd
64	172	5.0	104	1 LAC2_RAT	P20767 rattus norv
65	168.5	4.9	3707	1 PGBM_MOUSE	O05793 mus musculi
66	165.5	4.8	1493	1 NRC1_MOUSE	P97798 mus musculi
67	163	4.8	104	1 LAC3_MOUSE	P01845 mus musculi
68	162.5	4.8	213	1 IL11_HUMAN	P15814 homo sapien
69	162.5	4.8	847	1 CD22_HUMAN	P20273 homo sapien
70	161.5	4.7	739	1 VCA1_HUMAN	P19320 homo sapien
71	161	4.7	104	1 LAC2_MOUSE	P01844 mus musculi
72	161	4.7	258	1 HB2D_PIG	P15983 sus scrofa
73	161	4.7	702	1 CE45_HUMAN	P06731 homo sapien
74	159	4.7	105	1 LAC_HUMAN	P01842 homo sapien
75	156	4.6	105	1 LAC_PIG	P01846 sus scrofa
76	155	4.5	1906	1 KML5_CHICK	P11799 gallus gall
77	154	4.5	105	1 DTC_HUMAN	P01880 homo sapien
78	153.5	4.5	105	1 LAC5_MOUSEP	P20765 mus musculi
79	153	4.5	105	1 LAC_RABIT	P01847 oryctolagus
80	153	4.5	106	1 KACB_RABIT	P01839 oryctolagus
81	152	4.5	739	1 VCA1_MOUSE	P29533 mus musculi
82	150	4.4	106	1 KACA_RAT	P01836 rattus norv
83	150	4.4	509	1 SHS1_RAT	P97710 r protei
84	149	4.4	261	1 HB2C_PIG	P15982 sus scrofa
85	148.5	4.3	2012	1 DBCA_HUMAN	O60469 homo sapien
86	148	4.3	106	1 KAC_HUMAN	P01834 homo sapien
87	148	4.3	1447	1 DCC_MOUSE	P70211 mus musculi
88	147.5	4.3	503	1 SHS1_HUMAN	P78324 h protei
89	147	4.3	398	1 SRE1_HUMAN	O00241 homo sapien
90	147	4.3	6885	1 SNE2_HUMAN	O8wxn0 homo sapien
91	146	4.3	104	1 LAC1_RAT	P20766 rattus norv
92	146	4.3	387	1 SRE2_HUMAN	O91766 homo sapien
93	145.5	4.3	103	1 LAC_CHICK	P20763 gallus gall
94	145	4.2	106	1 KACB_RAT	P01835 rattus norv
95	144.5	4.2	555	1 C166_CARAU	O99304 caataxiu a
96	144.5	4.2	1257	1 CAML_HUMAN	P33004 homo sapien
97	144.5	4.2	1260	1 CAML_MOUSE	P11627 mus musculi
98	143	4.2	106	1 KAC_MOUSE	O31137 mus musculi
99	142	4.2	837	1 NCW2_MOUSE	P03592 homo sapien
100	141.5	4.1	761	1 CAML_HUMAN	P13592 homo sapien
101	141	4.1	1259	1 CAML_RAT	O05695 rattus norv
102	140	4.1	268	1 HB2X_HUMAN	P05338 homo sapien
103	140	4.1	273	1 DOB1_HUMAN	P13765 homo sapien
104	139.5	4.1	1234	1 NPBN_RAT	O90704 rattus norv
105	139	4.1	273	1 DOB1_PANTR	P18467 pan troglod
106	138.5	4.1	1197	1 CAM1_BRAAE	O90478 brachydantio

```

107 137.5 4.0 564 1 C166_BRARE Q09460 brachydanio
108 136 4.0 261 1 HB24_HUMAN P01920 homo sapien
109 136 4.0 997 1 SCPI_RAT O03410 rattus norv
110 135.5 4.0 105 1 LAC5_MOUSE P20764 mus musculu
111 135.5 4.0 848 1 NCAL_HUMAN P13591 homo sapien
112 135.5 4.0 1447 1 DCC_HUMAN P43594 homo sapien
113 134.5 3.9 261 1 HB22_HUMAN P01919 homo sapien
114 134.5 3.9 1348 1 VGR2_COTVA P52583 coturnix co
115 134 3.9 1369 1 NFAS_CHICK O42414 gallus galli
116 133.5 3.9 1242 1 NPHAS_MOUSE O9qz67 mus musculu
117 132.5 3.9 261 1 HB21_HUMAN P01918 homo sapien
118 132.5 3.9 725 1 NCAL_MOUSE P13594 mus musculu
119 132 3.9 231 1 HB2L_CHICK P23068 gallus galli
120 131.5 3.9 1240 1 NFAS_MOUSE O810u3 mus musculu
121 131 3.8 261 1 HB23_HUMAN P05537 homo sapien
122 130 3.8 837 1 NCM2_HUMAN O15394 homo sapien
123 130 3.8 1302 1 NRG_DROME P20241 drosophila
124 129.5 3.8 853 1 NCAL_BOVIN P1836 bos taurus
125 128 3.7 1097 1 PCDR_RAT Q05030 rattus norv

```

ALIGNMENTS

```

RESULT 1
CD4_HUMAN STANDARD: PRT; 458 AA.
AC P01730:

```

```

DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE CD4).
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=85254948; PubMed=2990730;
RA Maddon P.J., Littman D.R., Godfrey M., Maddon D.E., Chess L.,
RA Axel R.;
RT "The isolation and nucleotide sequence of a cDNA encoding the T cell
RT surface protein T4: a new member of the immunoglobulin gene family.";
RL Cell 42:93-104(1985).
RN [2]
RP REVISION TO 26.
RX MEDLINE=89028665; PubMed=3263213;
RA Littman D.R., Maddon P.J., Axel R.;
RT "Corrected CD4 sequence.";
RL Cell 55:541-541(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=96303695; PubMed=8723724;
RA Ananari-Lari M.A., Muzny D.M., Lu J., Lu F., Lilley C.E., Spanos S.,
RA Malley T., Gibbs R.A.;
RT "A gene-rich cluster between the CD4 and triosephosphate isomerase
RT genes at human chromosome 12p13.";
RL Genome Res. 6:314-326(1996).
RN [4]
RP SEQUENCE FROM N.A., AND VARIANT TRP-265.
RX MEDLINE=91216786; PubMed=1708753;
RA Hodge T.W., Sasse D.R., McDougal J.S.;
RT "Humans with OKT4 epitope deficiency have a single nucleotide base
RT change in the CD4 gene, resulting in substitution of TRP240 for
RT ARG240.";
RL Hum. Immunol. 30:99-104(1991).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas; PubMed=12477932;
RX MEDLINE=22386257;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

```

```

RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Bueto K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.T., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schreier T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Wooley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield Y.S.N., Krzywinski M.I., Skalka U., Smalins D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RP SEQUENCE OF 28-424 FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=93049640; PubMed=1425921;
RA Fomsgaard A., Hirsch V.M., Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
RT cellular receptor for simian immunodeficiency virus/human
RT immunodeficiency virus.";
RL Eur. J. Immunol. 22:2973-2981(1992).
RN [7]
RP SEQUENCE OF 26-394
RX MEDLINE=90078233; PubMed=2592374;
RA Carr S.A., Hemling M.E., Folena-Wasserman G., Sweet R.W., Anumula K.,
RA Barr J.R., Huddleston M.J., Taylor P.;
RT "Protein and carbohydrate structural analysis of a recombinant
RT soluble CD4 receptor by mass spectrometry.";
RL J. Biol. Chem. 264:21286-21295(1989).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 26-208.
RX MEDLINE=91061881; PubMed=1701030;
RA Wang J., Yan Y., Garrett T.P., Liu J., Rodgers D.W., Garlick R.L.,
RA Tarr G.E., Husain Y., Reinherz E.L., Harrison S.C.;
RT "Atomic structure of a fragment of human CD4 containing two
RT immunoglobulin-like domains.";
RL Nature 348:411-418(1990).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 26-208.
RX MEDLINE=91061882; PubMed=2247146;
RA Ryu S.-E., Kwong P.D., Truneh A., Porter T.G., Athos J.,
RA Rosenberg M., Dai X., Xuong N.-H., Axel R., Sweet R.W.,
RA Hendrickson W.A.;
RT "Crystal structure of an HIV-binding recombinant fragment of human
RT CD4.";
RL Nature 348:419-426(1990).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (3.9 ANGSTROMS) OF 26-388.
RX MEDLINE=97311402; PubMed=9168119;
RA Wu H., Kwong P.D., Hendrickson W.A.;
RT "Dimeric association and segmental variability in the structure of
RT human CD4.";
RL Nature 387:527-530(1997).
RN [11]
RP PALMITOYLATION.
RX MEDLINE=92317088; PubMed=1618861;
RA Crise B., Rose J.K.;
RT "Identification of palmitoylation sites on CD4, the human
RT immunodeficiency virus receptor.";
RL J. Biol. Chem. 267:13593-13597(1992).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.

```

CC	-I-	DATABASE:	NAME=PROV; NOTE=CD guide CD4 entry;
CC		WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd4.htm".	
CC	-----		
CC	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
CC	CC	the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way	
CC	CC	modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/announcements send an email to license@ebi.ac.uk).	
CC	CC	-----	
DR	EMBL:	M12807; AAA35572.1; --	
DR	EMBL:	U47924; AAB51309.1; --	
DR	EMBL:	M35160; AAAL6069.1; --	
DR	PIR:	BC025782; AAH25782.1; --	
DR	PDB:	1CDH; 30-APR-94.	
DR	PDB:	1CDI; 30-APR-94.	
DR	PDB:	3CDJ; 31-OCT-93.	
DR	PDB:	1CDU; 01-APR-97.	
DR	PDB:	1CDY; 01-APR-97.	
DR	PDB:	1WBR; 12-MAR-97.	
DR	PDB:	1WIO; 07-JUL-97.	
DR	PDB:	1WIP; 07-JUL-97.	
DR	PDB:	1WIQ; 07-JUL-97.	
DR	PDB:	1G9M; 27-DEC-00.	
DR	PDB:	1G9N; 27-DEC-00.	
DR	PDB:	1GC1; 19-AUG-98.	
DR	PDB:	1JL4; 19-SEP-01.	
DR	GlycoSuiteDB:	P01730; --	
DR	Genew:	HGNC:1678; CD4.	
DR	GO:	GO:0042101; C:T-cell receptor complex; NAS.	
DR	GO:	GO:0015026; F:coreceptor activity; NAS.	
DR	GO:	GO:0015029; F:internalization receptor activity; TAS.	
DR	GO:	GO:0004888; F:transmembrane receptor activity; TAS.	
DR	GO:	GO:0006955; P:immune response; NAS.	
DR	GO:	GO:0009405; P:pathogenesis; TAS.	
DR	GO:	GO:0030217; P:t-cell differentiation; NAS.	
DR	GO:	GO:0045058; P:t-cell selection; NAS.	
DR	InterPro:	IPRO00973; CD4 TCAG.	
DR	InterPro:	IPRO02110; IG-Like.	
DR	InterPro:	IPRO03596; Ig_V.	
DR	Pfam:	PF00047; Ig_2.	
DR	PRINTS:	PR00692; CD4TCANTIGEN.	
DR	SMART:	SMO0406; IGV_1.	
DR	PROSITE:	PS50835; IG_LIKE; 1.	
KW	Immunoglobulin domain,	Transmembrane; Glycoprotein; T-cell;	
KW	Immune response; Repeat;	Signal; Lipoprotein; Palmitate; 3D-structure;	
KW	Polymorphism.		
FT	SIGNAL	1 25	
FT	CHAIN	26 458	
FT	DOMAIN	26 396	
FT	TRANSMEM	397 418	
FT	DOMAIN	419 458	
FT	DOMAIN	26 125	
FT	DOMAIN	126 203	
FT	DOMAIN	204 317	
FT	DOMAIN	318 374	
FT	CARBOHYD	296	
FT			
FT	CARBOHYD	325 325	
FT			
FT	DISULFID	41 109	
FT	DISULFID	155 184	
FT	LIPID	328 370	
FT	LIPID	419 419	
FT	LIPID	422 422	
FT	VARIANT	265 265	
FT			
FT	R -> W	(in OKT4-negative populations) .	

FT	STRAND	27	32		/FTID=VAR_003906.
FT	TUN	33	34		
FT	STRAND	37	39		
FT	TUN	44	45		
FT	STRAND	51	55		
FT	TUN	56	57		
Query Match 59.3%; Score 2024; DB 1; Length 458;					
Best Local Similarity 99.2%; Pred. No. 1,4e-122;					
Matches 393; Conservative 1; Mismatches 2; Indels 0; Gaps 0;					
QY	1	MNRGVPFRHLLVLTQALLPAAATGKNVYLGKKDYTELTCTASQKKSIOFHMKNNOIK	60		
DB	1	MNRGVPFRHLLVLTQALLPAAATGKNVYLGKKDYTELTCTASQKKSIOFHMKNNOIK	60		
QY	61	ILNGQSFLLTKPGSKLMDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEVQL	120		
DB	61	ILNGQSFLLTKPGSKLMDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEVQL	120		
QY	121	LVFELTANSDPHLLQGGSLTLETSPGSSPSVQCRPRGNIGQKTLVSQLELDQSG	180		
DB	121	LVFELTANSDPHLLQGGSLTLETSPGSSPSVQCRPRGNIGQKTLVSQLELDQSG	180		
QY	181	TWCTTVONOKKVEFKDIDIVLAFQKASSIYKKEGQVEFSPLAFTVEKLTGSGELMW	240		
DB	181	TWCTTVONOKKVEFKDIDIVLAFQKASSIYKKEGQVEFSPLAFTVEKLTGSGELMW	240		
QY	241	QAEBASSKSWITFDLKNKEVSVKRVYQDPKLQWKKLPLHLTLPOALPOYAGSGNLTLLA	300		
DB	241	QAEBASSKSWITFDLKNKEVSVKRVYQDPKLQWKKLPLHLTLPOALPOYAGSGNLTLLA	300		
QY	301	LEATGKLGHEVNLVYWRATQLOKMLTCEVWGPTSPKLMSTLKENKBAKSKREKPVVY	360		
DB	301	LEATGKLGHEVNLVYWRATQLOKMLTCEVWGPTSPKLMSTLKENKBAKSKREKPVVY	360		
QY	361	LNPEAGMWOCCLSDSGOVLLESNTKVLPTWSTPVEP	396		
DB	361	LNPEAGMWOCCLSDSGOVLLESNTKVLPTWSTPVEP	396		
RESULT 2					
CD4_PANTR	STANDARD;	PRT;	458 AA.		
AC	P16004;				
DT	01-APR-1990 (Rel. 14, Created)				
DT	01-APR-1990 (Rel. 14, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	T-cell surface glycoprotein CD4 precursor (T-cell surface antigen CD4.				
OS	Pan troglodytes (Chimpanzee).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.				
OX	NCBI_TaxID=9598;				
RN	[1]				
RN	SEQUENCE FROM N.A.				
RX	MEDLINE=90182664; PubMed=2107024;				
RA	Camerini D., Seed B.;				
RT	"A CD4 domain important for HIV-mediated syncytium formation lies outside the virus binding site."				
RT	Cell 60:747-754(1990).				
RN	[2]				
RP	SEQUENCE OF 26-424 FROM N.A.				
RC	TISSUE=Blood;				
RX	MEDLINE=93049640; PubMed=1425521;				
RA	Fomsgaard A., Hirsch V.M., Johnson P.R.;				
RT	"Cloning and sequences of primate CD4 molecules: diversity of the cellular receptor for simian immunodeficiency virus/human immunodeficiency virus."				
RT	Eur. J. Immunol. 22:2973-2981(1992).				
CC	-1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell receptor interaction. May regulate T-cell activation.				

```

CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M31135; AAA5407.1; -.
DR EMBL; X73323; CAA51749.1; -.
DR PIR; B32722; RMCZT4.
DR HSSP; P01730; IMIO.
DR GO; GO:0042101; C: T-cell receptor complex; ISS.
DR GO; GO:0015026; F: coreceptor activity; ISS.
DR GO; GO:0042289; F: MHC class II protein binding; ISS.
DR GO; GO:0006955; P: immune response; ISS.
DR GO; GO:0045086; P: positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P: T-cell differentiation; ISS.
DR GO; GO:0045058; P: T-cell selection; ISS.
DR GO; GO:0007169; P: transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396
FT TRANSMEM 397 418
FT DOMAIN 419 458
FT DOMAIN 126 125
FT DOMAIN 126 203
FT DOMAIN 204 317
FT CARBOHYD 296 374
FT CARBOHYD 325 325
FT DISULFID 41 109
FT DISULFID 155 184
FT DISULFID 328 370
FT LIPID 419 419
FT LIPID 422 422
FT CONFLICT 42 62
FT CONFLICT 62 62
FT CONFLICT 191 191
SQ SEQUENCE 458 AA; 51057 MW; A7C3AC8A5257D3AD CRC64;
-----
Query Match 58.6%; Score 2000; DB 1; Length 458;
Best Local Similarity 98.2%; Pred. No. 4-7e-12;
Matches 389; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

```

```

Db 181 TWCTCTVLONOKKVEFKIDIVLAFQKASIVYKKEGQVEFSPPLAFTVEKLTSSGELMW 240
Oy QAERASSKSMITFDLKNKEVSVKRVTPDRLQWGGKLLPHLTLPALPOYAGSGNTTLA 300
Db 241 QAERASSKSMITFDLKNKEVSVKRVTPDRLQWGGKLLPHLTLPALPOYAGSGNTTLA 300
Oy 301 LEAKTGKHOEVLVVRATOLQKNLTCEVWGPTSPKMLSLKLENKARVSKKEPVW 360
Db 301 LEAKTGKHOEVLVVRATOLQKNLTCEVWGPTSPKMLSLKLENKARVSKKEPVW 360
Oy 361 LNPEAGWOCLLSDSGVLLSESNIKVLPTWSTPVP 396
Db 361 LNPEAGWOCLLSDSGVLLSESNIKVLPTWSTPVP 396
-----
RESULT 3
CD4_MACFU STANDARD; PRT; 458 AA.
ID CD4_MACFU
AC P79184;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3).
GN CD4.
OS Macaca fuscata fuscata (Japanese macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_Taxid=9543;
RN [1]
RP SEQUENCE FROM N.A.
RA Hashimoto O., Tatsumi M.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; D63348; BAA09672.1; -.
DR HSSP; P01730; IMBR.
DR GO; GO:0042101; C: T-cell receptor complex; ISS.
DR GO; GO:0015026; F: coreceptor activity; ISS.
DR GO; GO:0042289; F: MHC class II protein binding; ISS.
DR GO; GO:0006955; P: immune response; ISS.
DR GO; GO:0045086; P: positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P: T-cell differentiation; ISS.
DR GO; GO:0045058; P: T-cell selection; ISS.
DR GO; GO:0007169; P: transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396
FT TRANSMEM 397 418
FT DOMAIN 419 458
SQ SEQUENCE 458 AA; 51057 MW; A7C3AC8A5257D3AD CRC64;

```

FT	DOMAIN	26	125	IG-LIKE V-TYPE.
FT <td>DOMAIN</td> <td>126</td> <td>203</td> <td>IG-LIKE C2-TYPE 1.</td>	DOMAIN	126	203	IG-LIKE C2-TYPE 1.
FT <td>DOMAIN</td> <td>204</td> <td>317</td> <td>IG-LIKE C2-TYPE 2.</td>	DOMAIN	204	317	IG-LIKE C2-TYPE 2.
FT <td>DOMAIN</td> <td>318</td> <td>374</td> <td>IG-LIKE C2-TYPE 3.</td>	DOMAIN	318	374	IG-LIKE C2-TYPE 3.
FT <td>CARBOHYD</td> <td>42</td> <td>42</td> <td>(POTENTIAL).</td>	CARBOHYD	42	42	(POTENTIAL).
FT <td>CARBOHYD</td> <td>286</td> <td>286</td> <td>(BY SIMILARITY).</td>	CARBOHYD	286	286	(BY SIMILARITY).
FT <td>CARBOHYD</td> <td>325</td> <td>325</td> <td>(BY SIMILARITY).</td>	CARBOHYD	325	325	(BY SIMILARITY).
FT <td>DISULFID</td> <td>41</td> <td>109</td> <td>(BY SIMILARITY).</td>	DISULFID	41	109	(BY SIMILARITY).
FT <td>DISULFID</td> <td>155</td> <td>184</td> <td>(BY SIMILARITY).</td>	DISULFID	155	184	(BY SIMILARITY).
FT <td>DISULFID</td> <td>328</td> <td>370</td> <td>(BY SIMILARITY).</td>	DISULFID	328	370	(BY SIMILARITY).
FT <td>LIPID</td> <td>419</td> <td>419</td> <td>(BY SIMILARITY).</td>	LIPID	419	419	(BY SIMILARITY).
FT <td>LIPID</td> <td>422</td> <td>422</td> <td>(BY SIMILARITY).</td>	LIPID	422	422	(BY SIMILARITY).
SO <td>SEQUENCE</td> <td>458 AA.</td> <td>50828 MW;</td> <td>7683E7F08185535 CR664;</td>	SEQUENCE	458 AA.	50828 MW;	7683E7F08185535 CR664;
Query Match 54.3%; Score 1853; DB 1; Length 458;				
Best Local Similarity 90.2%; Pred. No. 1.3e-11;				
Matches 357; Conservative 18; Mismatches 21; Indels 0; Gaps 0				
QY	1	MNRGVPFPHLLVLTOLALIPATQGNKVVLGKKGDTVELCTASQKKSIOFHWKSNQIK	60	
DB	1	MNRGIPFPHLLVLTOLALIPATQGNKVVLGKKGDTVELCTASQKKSIOFHWKSNQIK	60	
QY	61	ILGNQSPFLTKPSKLNPDARSRLMDQGNPLIIKNIKINDSPYICEVDQKEEYOL	120	
DB	61	ILGIGSPFLTKPSKLNPDARSRLMDQGCFSMIINKIKINDSPYICEVENKKEEVEL	120	
QY	121	LVFGLTANSDFLLLOGOSLTLTLFLESPPGSSPVSQCRSPGKNIQGGKTLVSQLELDQSG	180	
DB	121	LVFGLTANSDFLLLEGOSLTLTLFLESPPGSSPVSQCRSPGKNIQGGRTISVQLERQDSG	180	
QY	181	TWTCTVLQNOKKVEFKIDIVLAFQKASSIVYKKSGEYFSPPLAFVTEKLTGSGELMW	240	
DB	181	TWTCTVSODQKVEFKIDIVLAFQKASSIVYKKEGEYFSPPLAFLEKLTGSGELMW	240	
QY	241	QAEBSASSKSWTTFPLKKNKESVSKVTQOPKQMGKPLPHLTLPQALQVYAGSGNLTLA	300	
DB	241	QAEBSASSKSWTTFPLKKNKESVSKVTQOPKQMGKPLPHLTLPQALQVYAGSGNLTLA	300	
QY	301	LEAKTGKLGHOENLVLMRATQLOKQLTCEVMGPTSPKMLSLKLEKAKVSKREKPVNV	360	
DB	301	LEAKTGKLGHOENLVLMRPAQOENLTCVMEPTSPKLTLSKLEKAKVSKQAKAVNV	360	
QY	361	LNPEAGMWQCLLSDSGOVLLESINIKVLPWTSTPVEZ	396	
DB	361	LNPEAGMWQCLLSDSGOVLLESINIKVLPWTPVPVQ	396	
RESULT 4				
ID	CD4_MACMU	STANDARD;	PRT;	458 AA.
AC	P16003; Q29617;			
DT	01-APR-1990 (Rel. 14, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	T-cell surface glycoprotein CD4 precursor (T-cell surface antigen			
DE	T4/Leu-3).			
GN	CD4.			
OS	Macaca mulatta (rhesus macaque).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;			
OC	Cercopithecoidea; Macaca.			
OX	NCBI_TaxID=9544;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=90182664; PubMed=2107024;			
RA	Camerini D., Seed B.;			
RT	"A CD4 domain important for HIV-mediated syncytium formation lies			
RT	outside the virus binding site.";			
RL	Cell 60:747-754(1990).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Thymocytes;			
RA	Haehimoto O., Tatsumi M.;			

RT	"Molecular cloning and expression of maeaque CD4c.";
RL	Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN	[3]
RP	SEQUENCE OF 28-424 FROM N.A.
RC	TISSUE=Blood;
RX	MEDLINE=93049640; PubMed=1425921;
RA	Pomgaard A., Hirsch V.M., Johnson P.R.;
RT	"Cloning and sequences of primate CD4 molecules: diversity of the cellular receptor for simian immunodeficiency virus/human immunodeficiency virus."/;
RL	Eur. J. Immunol. 22:2973-2981(1992).
RN	[4]
RP	SEQUENCE OF 107-192 FROM N.A.
RX	MEDLINE=98320644; PubMed=9656488;
RA	Harris E.E., Disotell T.R.;
RT	"Nuclear gene trees and the phylogenetic relationships of the mangabays (Primates: Papionini).";
RL	Mol. Biol. Evol. 15:892-900(1998).
CC	-I- FUNCTION: Accessory protein for MHC class-II antigen/T-cell receptor interaction. May regulate T-cell activation.
CC	-I- SUBUNIT: Associates with p56-lck (By similarity).
CC	-I- SUCCELLULAR LOCATION: Type I membrane protein.
CC	-I- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC	-I- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL: M31134; AAA56838.1; -
DR	EMBL: D63347; BAA09671.1; -
DR	EMBL: X73326; CAAS1752.1; -
DR	EMBL: AF057385; AAC25129.1; -
DR	HSSP: P01730; IMBR.
DR	GO: GO:0042101; C:T-cell receptor complex; ISS.
DR	GO: GO:0015026; F:coreceptor activity; ISS.
DR	GO: GO:0042289; F:MHC class II protein binding; ISS.
DR	GO: GO:0006959; F:Immune response; ISS.
DR	GO: GO:0045086; P:Positive regulation of interleukin-2 biosyn. . . ; ISS.
DR	GO: GO:0030217; P:T-cell differentiation; ISS.
DR	GO: GO:0045058; P:T-cell selection; ISS.
DR	GO: GO:0007169; P:transmembrane receptor protein tyrosine kin. . . ; ISS.
DR	InterPro: IPR000973; CD4 TCAG.
DR	InterPro: IPR007110; Ig_Like.
DR	InterPro: IPR003596; Ig_V.
DR	pfam: PF00047; Ig_2.
DR	PRINTS: PR006592; CD4TCANTIGEN.
DR	SMART: SM00406; IGV_1
KW	PROSITE: PS50835; IG_LIKE; 1.
KW	Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW	Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT	SIGNAL 1 25
FT	CHAIN 26 458
FT	DOMAIN 26 396
FT	TRANSMEM 397 418
FT	DOMAIN 419 458
FT	DOMAIN 26 125
FT	DOMAIN 126 203
FT	DOMAIN 204 317
FT	DOMAIN 318 374
FT	CARBOHYD 296 296
FT	CARBOHYD 325 325
FT	DISULFID 41 109
FT	DISULFID 155 184
FT	DISULFID 328 370
FT	LIPID 419 419
FT	LIPID 422 422
FT	CONFLICT 42 42
FT	CONFLICT 62 62
FT	L -> S (IN REF. 3).

```

FT CONFLICT 67 67 L -> S (IN REF. 2).
FT CONFLICT 169 169 I -> L (IN REF. 2).
FT CONFLICT 191 191 K -> N (IN REF. 3).
FT CONFLICT 248 248 S -> P (IN REF. 2).
FT CONFLICT 265 265 R -> Q (IN REF. 3).
FT CONFLICT 349 349 A -> T (IN REF. 2).
SQ SEQUENCE 458 AA; 50884 MW; 88880339FAFE808 CRC64;

Query Match 54.2%; Score 1852; DB 1; Length 458;
Best Local Similarity 90.2%; Pred. No. 1.5e-11;
Matches 357; Conservative 18; Mismatches 21; Indels 0; Gaps 0;

OY 1 MNRGVPFRHLVLVQLALPPATQGNKVLGKGGDTVELTCTASQKSIQFHMKNQIK 60
DB 1 MNRGIPFRHLVLVQLALPPATQGNKVLGKGGDTVELTCTASQKSIQFHMKNQIK 60
OY 61 ILNGSFLTKGSPSKLNDRAISRRLMDOGNFPLIINKLTIEDSDTYICEVEDQKEVQL 120
DB 61 ILGIGFLTKGSPSKLNDRAISRRLMDOGNFPLIINKLTIEDSDTYICEVENKKEVEL 120
OY 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDSG 180
DB 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDSG 180
OY 181 TWCTCTVONOKKVEFKIDIVLAFQKASSIYKKEGQVFSFPLATFVEKLTGSGELMW 240
DB 181 TWCTCTVONOKKVEFKIDIVLAFQKASSIYKKEGQVFSFPLATFVEKLTGSGELMW 240
OY 241 QABRASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPHLTLTLPQALPYAGSGNLTLLA 300
DB 241 QABRASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPHLTLTLPQALPYAGSGNLTLLA 300
OY 301 LEAKTGKHOEVNLVVMRATQKNTLCEVWGPTSPKMLSLKLENKAVSKREKPYVW 360
DB 301 LEAKTGKHOEVNLVVMRATQKNTLCEVWGPTSPKMLSLKLENKATVSKAKAVW 360
OY 361 LNPEAGMWQCLSDSGVLLSNIKYVLPWTSTPEP 396
DB 361 LNPEAGMWQCLSDSGVLLSNIKYVLPWTSTPEP 396

RESULT 5
ID CD4_MACFA STANDARD; PRT; 458 AA.
AC P79185;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen T4/Lew-3).
GN DN
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OC NCBI_TaxID=9541;
OX [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Thymocytes;
RC Tarsus M., Yabe M., Yamada Y.K.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D63349; BAA09673.1; -.
DR HSSP: P01730; IWR.
DR GO: GO:0042101; C: T-cell receptor complex; ISS.
DR GO: GO:0015026; F: coreceptor activity; ISS.
DR GO: GO:0042289; F: MHC class II protein binding; ISS.
DR GO: GO:0006955; P: immune response; ISS.
DR GO: GO:0045086; P: positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO: GO:0030217; P: T-cell differentiation; ISS.
DR GO: GO:0045058; P: T-cell selection; ISS.
DR GO: GO:0007169; P: transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro: IPR000973; CD4_TcRg.
DR InterPro: IPR007110; Ig_Like.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig_2.
DR PRINTS: PR00692; CD4TCRNTIGEN.
DR SMART: SM00406; IG_V_1.
DR PROSITE: PS50835; IG_LIKE_1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396 T-CELL SURFACE GLYCOPROTEIN CD4.
FT TRANSLEM 397 418 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 419 458 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 26 125 IG-LIKE V-TYPE.
FT DOMAIN 126 203 IG-LIKE C2-TYPE 1.
FT DOMAIN 204 317 IG-LIKE C2-TYPE 2.
FT DOMAIN 318 374 IG-LIKE C2-TYPE 3.
FT CARBOHYD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT DISULFID 41 109 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT DISULFID 155 184 BY SIMILARITY.
FT DISULFID 328 370 BY SIMILARITY.
FT LIPID 419 419 S-palmitoyl cysteine (By similarity).
FT LIPID 422 422 S-palmitoyl cysteine (By similarity).
SQ SEQUENCE 458 AA; 50872 MW; 9105479FB5C56FE7 CRC64;

Query Match 54.2%; Score 1849; DB 1; Length 458;
Best Local Similarity 90.2%; Pred. No. 2.3e-11;
Matches 357; Conservative 17; Mismatches 22; Indels 0; Gaps 0;

OY 1 MNRGVPFRHLVLVQLALPPATQGNKVLGKGGDTVELTCTASQKSIQFHMKNQIK 60
DB 1 MNRGIPFRHLVLVQLALPPATQGNKVLGKGGDTVELTCTASQKSIQFHMKNQIK 60
OY 61 ILNGSFLTKGSPSKLNDRAISRRLMDOGNFPLIINKLTIEDSDTYICEVEDQKEVQL 120
DB 61 ILGIGFLTKGSPSKLNDRAISRRLMDOGNFPLIINKLTIEDSDTYICEVENKKEVEL 120
OY 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDSG 180
DB 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDSG 180
OY 181 TWCTCTVONOKKVEFKIDIVLAFQKASSIYKKEGQVFSFPLATFVEKLTGSGELMW 240
DB 181 TWCTCTVONOKKVEFKIDIVLAFQKASSIYKKEGQVFSFPLATFVEKLTGSGELMW 240
OY 241 QABRASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPHLTLTLPQALPYAGSGNLTLLA 300
DB 241 QABRASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPHLTLTLPQALPYAGSGNLTLLA 300
OY 301 LEAKTGKHOEVNLVVMRATQKNTLCEVWGPTSPKMLSLKLENKAVSKREKPYVW 360
DB 301 LEAKTGKHOEVNLVVMRATQKNTLCEVWGPTSPKMLSLKLENKATVSKAKAVW 360
OY 361 LNPEAGMWQCLSDSGVLLSNIKYVLPWTSTPEP 396
DB 361 LNPEAGMWQCLSDSGVLLSNIKYVLPWTSTPEP 396

```

```

RESULT 6
CD4_MACNE
ID CD4_MACNE STANDARD; PRT; 458 AA.
AC 008340; P79196;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
T4/Lew-3).
GN CD4.
OS Macaca nemestrina (Pig-tailed macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OC NCBI_TaxID=9545;
RX [1]
RP SEQUENCE FROM N.A.
RA Hashimoto O., Tateumi M.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 28-424 FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=93049640; PubMed=1425921;
RX Fomsgaard A., Hirsch V.M., Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
RT cellular receptor for simian immunodeficiency virus/human
RT immunodeficiency virus."
RL Eur. J. Immunol. 22:2973-2981(1992).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; D63346; BA09670.1; -
DR HSSP; X73325; CAA51751.1; -
DR HSSP; P01730; 1MR.
DR GO; GO:0042101; C:T-cell receptor complex; ISS.
DR GO; GO:0015026; P:coreceptor activity; ISS.
DR GO; GO:0042289; P:MHC class II protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. .; ISS.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IGV_Like; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 397 418
FT DOMAIN 419 458 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 26 125
FT DOMAIN 126 203 IG-LIKE C2-TYPE 1.
FT DOMAIN 204 317 IG-LIKE C2-TYPE 2.
FT DOMAIN 318 374 IG-LIKE C2-TYPE 3.

```

```

FT CARBOHYD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT CARBOHYD 325 325 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT DISULFID 41 109 BY SIMILARITY.
FT DISULFID 155 184 BY SIMILARITY.
FT DISULFID 328 370 BY SIMILARITY.
FT LIPID 419 419 S-palmitoyl cysteine (By similarity).
FT LIPID 422 422 S-palmitoyl cysteine (By similarity).
FT CONFLICT 57 57 D -> N (IN REF. 2).
FT CONFLICT 91 91 C -> H (IN REF. 2).
FT CONFLICT 105 105 N -> D (IN REF. 2).
FT CONFLICT 113 113 N -> D (IN REF. 2).
FT CONFLICT 302 302 D -> E (IN REF. 2).
FT CONFLICT 349 349 T -> A (IN REF. 2).
SQ SEQUENCE 458 AA; 50905 MW; 751A9BA2C8B3E16 CRC64;

Query Match 54.0%; Score 1844; DB 1; Length 458;
Best Local Similarity 89.6%; Pred. No. 4,7e-111;
Matches 355; Conservative 20; Mismatches 21; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLIVLQALIPATQGNKVLGKKDVTLELTGASQKSIQPHMKNSQIK 60
DB 1 MNRGIPFRHLIVLQALIPAVTQKKVVLGKKDVTLELTGASQKKTQPHMKNSDQIK 60
QY 61 ILGNQGSFETKPSKLANRADSRSLMPDQNPPLIKKLTEDSDTYICEVDQEEYQL 120
DB 61 ILGIGSFETKPSKLSBRADSRKSLMDQCSMTIKKLTEDSDTYICEVNEKEVEYL 120
QY 121 LVFGLTANSDFHLLOGQSITLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDFHLBGQSITLTLESPPSSPSVQCRSPRKNIOGKTLVSQLELDQSG 180
QY 181 TWTCTVLQNKQKVEKIDIVLAFQKASSIYKKEGEVSEFPAFTVEKLTSGGELMW 240
DB 181 TWTCTVSDQKVEKIDIVLAFQKASSIYKKEGEVSEFPAFTVEKLTSGGELMW 240
QY 241 QERASSSKSWITPFLKNKEVSVKRVTPDPLQWKKPLPHLTLPQALPOVAGSGLTLTA 300
DB 241 QERASSSKSWITPFLKNKEVSVKRVTPDPLQWKKPLPHLTLPQALPOVAGSGLTLTA 300
QY 301 LEAKTKGLHGVNVLVMBATOLQKXLTCEVWGPTSPKMLSLKLNKAKVSKREKPYWV 360
DB 301 LDKAKTKGLHGVNVLVMBATOLQKXLTCEVWGPTSPKMLSLKLNKAKVSKREKPYWV 360
QY 361 LNPEAGMOCILSDSGQVLTLESNTIVLPTWSTPVP 396
DB 361 LNPEAGMOCILSDSGQVLTLESNTIVLPTWSTPVP 396

RESULT 7
CD4_CERAE
ID CD4_CERAE STANDARD; PRT; 458 AA.
AC 008338; 002805; 077593; Q28217;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
T4/Lew-3).
GN CD4.
OS Cercopithecus aethiops (Green monkey) (Grivet).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OC NCBI_TaxID=9534;
RX [1]
RP SEQUENCE FROM N.A.
RA Hashimoto O., Tateumi M.;
RL "Molecular cloning and expression of african green monkey CD4.";
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 28-424 FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=93049640; PubMed=1425921;

```

RA Fomsgaard A., Hirsch V.M., Johnson P.R.;
 RT "Cloning and sequences of primate CD4 molecules: diversity of the
 RT cellular receptor for simian immunodeficiency virus/human
 RT immunodeficiency virus.";
 RL Eur. J. Immunol. 22:2973-2981(1992).
 RN [3]
 RP SEQUENCE OF 28-424 FROM N.A.
 RC TISSUE=peripheral blood;
 RX Fomsgaard A., Mueller-Trutwin M.C., Dlop O., Hansen J., Mathiot C.,
 RA Corbet S., Barre-Sinoussi F., Allan J.S.;
 RT "Relation between phylogeny of African green monkey CD4 genes and
 RT their respective simian immunodeficiency virus genes.";
 RL J. Med. Primatol. 26:120-128(1997).
 RN [4]
 RP SEQUENCE OF 107-192 FROM N.A.
 RX MEDLINE=98320644; PubMed=9656488;
 RA Harris E.E., Disotell T.R.;
 RT "Nuclear gene trees and the phylogenetic relationships of the
 RT mangabays (Primates: Papionini).";
 RL Mol. Biol. Evol. 15:892-900(1998).
 CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
 CC receptor interaction. May regulate T-cell activation.
 CC -1- SUBUNIT: Associates with p56^{lck} (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, D86589, BAAL332.1, -;
 DR EMBL, X73322, CAAS1748.1, -;
 DR EMBL, AF001226, AAB60873.1, -;
 DR EMBL, AF001228, AAB60875.1, -;
 DR EMBL, AF057380, AAC25124.1, -;
 DR HSRP, P01730, IWO,
 DR GO, GO:0042101, C-T-cell receptor complex; ISS.
 DR GO, GO:0015026, F:coreceptor activity; ISS.
 DR GO, GO:0042289, F:MHC class II protein binding; ISS.
 DR GO, GO:0006955, P:immune response; ISS.
 DR GO, GO:0045086, P:positive regulation of interleukin-2 biosyn. . .; ISS.
 DR GO, GO:0030217, P:T-cell differentiation; ISS.
 DR GO, GO:0045058, P:T-cell selection; ISS.
 DR GO, GO:0007169, P:transmembrane receptor protein tyrosine kin. . .; ISS.
 DR InterPro: IPR000973; CD4 TCAG.
 DR InterPro: IPR007110; IG-like.
 DR InterPro: IPR003596; IG_v.
 DR Pfam: PF00047; Ig_2.
 DR PRINTS, PRO0692; CD4TCANTIGEN.
 DR SMART, SM00406; IGV; 1.
 DR PROSITE, PS50835; IG_LIKE; 1.
 DR Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
 KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
 FT SIGNAL 1 25
 FT CHAIN 26 458
 FT DOMAIN 26 396
 FT TRANSMEM 397 418
 FT DOMAIN 419 458
 FT DOMAIN 26 125
 FT DOMAIN 204 317
 FT DOMAIN 318 374
 FT CARBOHYD 42 42
 FT CARBOHYD 281 281
 FT CARBOHYD 296 296
 FT CARBOHYD 325 325
 FT DISULFID 41 109
 BY SIMILARITY.

FT DISULFID 155 184 BY SIMILARITY.
 FT DISULFID 318 370 BY SIMILARITY.
 FT LIPID 429 439 S-palmitoyl cysteine (By similarity).
 FT LIPID 422 422 S-palmitoyl cysteine (By similarity).
 FT CONFLICT 46 46 K -> N (IN REF. 2 AND 3; AAB60875).
 FT CONFLICT 59 59 I -> T (IN REF. 3; AAB60873).
 FT CONFLICT 115 115 K -> E (IN REF. 1).
 FT CONFLICT 165 165 G -> V (IN REF. 3; AAB60873 AND 4).
 FT CONFLICT 200 200 M -> L (IN REF. 2 AND 3).
 FT CONFLICT 227 227 F -> L (IN REF. 3; AAB60873).
 FT CONFLICT 271 271 K -> E (IN REF. 3; AAB60873).
 FT CONFLICT 281 281 N -> H (IN REF. 3; AAB60873).
 SO SEQUENCE 458 AA, 51158 MW, FC523D2EDDLF72E7 CRC64;
 Query Match 53.5%; Score 1827; DB 1; Length 458;
 Best Local Similarity 89.1%; Pred. No. 5.8e-110;
 Matches 355; Conservative 18; Mismatches 25; Indels 0; Gaps 0;
 QY 1 MNRGVPFRHLLVLVQLALPAAATQGNKRVLGKGDVLELTCTASQKKSIOFHKNSNQIK 60
 DB 1 MNMGIPFRHLLVLVQLALPAAVTOGKKVVGKGDVLELTCTASQKKTTFPHMKNQIK 60
 QY 61 ILGQGSFLTKGSPSKNDRAISRSLMDGNFPLIKNLKIEPDTYICFVEDQKERVOL 120
 DB 61 ILGQGSFLTKGSKLDRIDRSKSLMDQCFSHILNLKIEBETIICEVENKKEVEL 120
 QY 121 LVFGLTANSDTHLLOGSLTLTLESPGSSPSVQCRSPRKNIOGKTLVSQLELQDSG 180
 DB 121 LVFGLTANSDTHLLOGSLTLTLESPGSSPSVQCRSPRKNIOGKTLVSQLELQDSG 180
 QY 181 TWTTTVLQNKVVEFKIDIVLAFQKASSIVYKKEGQVEPSPPLATVEKLTSGGLMW 240
 DB 181 TWTTTVLQNKVVEFKIDIVLAFQKASSIVYKKEGQVEPSPPLATVEKLTSGGLMW 240
 QY 241 QABRASSKSMITFDLKNKEVSVKRVTOPDKLQMGKPLPHLTLPOLPOYAGSGNTLA 300
 DB 241 QABRASSKSMITFDLKNKEVSVKRVTOPDKLQMGKPLPHLTLPOLPOYAGSGNTLA 300
 QY 301 LEATGKLHOEVNLVVRATOLQKNLTCEVWGPTSPKMLSTLKNKAAYSKREKPVW 360
 DB 301 LEATGKLHOEVNLVVRATOLQKNLTCEVWGPTSPKMLSTLKNKAAYSKREKPVW 360
 QY 361 LNPEAGWQCLSDSGQVLLSNIKVLPTWSTPYEP 396
 DB 361 LNPEAGWQCLSDSGQVLLSNIKVLPTWSTPYEP 396
 RESULT 8
 CD4_ERYPA STANDARD; PRT; 397 AA.
 ID CD4_ERYPA
 AC Q08339;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/Leu-3)
 DE (Fragment).
 OS CD4.
 OS Erythrocybus pates (Red quonon) (Huesar).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Erythrocybus.
 OC NCB1 TaxID=9538;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Blood.
 RX MEDLINE=93049640; PubMed=1425921;
 RA Fomsgaard A., Hirsch V.M., Johnson P.R.;
 RT "Cloning and sequences of primate CD4 molecules: diversity of the
 RT cellular receptor for simian immunodeficiency virus/human
 RT immunodeficiency virus.";
 RL Eur. J. Immunol. 22:2973-2981(1992).
 CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
 CC receptor interaction. May regulate T-cell activation.


```

CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC the SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X73324; CA51750.1; -.
DR HSBP; P01730; 1W1Q.
DR GO: GO:0042101; C: T-cell receptor complex; ISS.
DR GO: GO:0015026; F: coreceptor activity; ISS.
DR GO: GO:0042289; F: MHC class II protein binding; ISS.
DR GO: GO:0006955; P: immune response; ISS.
DR GO: GO:0045086; P: positive regulation of interleukin-2 biosyn. .; ISS.
DR GO: GO:0030211; P: T-cell differentiation; ISS.
DR GO: GO:0007169; P: transmembrane receptor protein tyrosine kin. .; ISS.
DR InterPro: IPR0003973; C4A TCAG.
DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR003596; Ig_LV.
DR Pfam; PF00047; Ig; 2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG-LIKE; 1.
KW Immunoglobulin domain, Transmembrane, Glycoprotein, T-cell;
KW Immune response; Repeat; Lipoprotein; Palmitate.
KM
FT NON_TER 1 1
FT DOMAIN <1 369 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 370 391 POTENTIAL.
FT DOMAIN 392 >397 CYTOPLASMIC (POTENTIAL).
FT DOMAIN <1 98 IG-LIKE V-TYPE.
FT DOMAIN 99 176 IG-LIKE C2-TYPE 1.
FT DOMAIN 177 290 IG-LIKE C2-TYPE 2.
FT DOMAIN 291 343 IG-LIKE C2-TYPE 3.
FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT DISULFID 14 82 BY SIMILARITY.
FT DISULFID 128 157 BY SIMILARITY.
FT DISULFID 301 343 S-palmitoyl cysteine (By similarity).
FT LIPID 392 392 S-palmitoyl cysteine (By similarity).
FT LIPID 395 395 S-palmitoyl cysteine (By similarity).
FT NON_TER 397 397
SO SEQUENCE 397 AA; 44081 MW; 678873976AB7E4AF CRC64;

Query Match 50.6%; Score 1726; DB 1; Length 397;
Best Local Similarity 89.7%; Pred. No. 1,4e-103;
Matches 331; Conservative 16; Mismatches 22; Indels 0; Gaps 0;

QY 28 VLKGGDTVELTCTAAGKKSIOFHWNKSNQIKLGNQGSFLYTGPSKLNDRADSRSLW 87
DB 1 VLKGGKGTVELTCTAASQKTTQFHWKNSNMOKILGQGSFLYTGPSKLNDRADSRKSLW 60
QY DQGNRPLLIKKLIKEDSDTYICEVNDQKEEYQVLVFGLTANSDPHLQGSSTLTLESPP 147
DB 1 DQGCFSMIIKKLIKEDSETYICEVDKKEEVELLVFGLTANSDPHLQGSSTLTLESPP 120
QY 148 GSSPSVQCRSPRGKNIQGGKTLVSQQLAOSGTCTVTLNOKKVERKIDIVLAFOKA 207
DB 121 GSSPSVQCRSPRGKNIQGGKTLVSQQLAOSGTCTVTLNOKKVERKIDIVLAFOKA 180
QY 208 SSIYKKEGEVSEFSLAFLVEKLTSGGELIWMQERASSSSKSWITPDLKKEVSVKRYT 267
DB 181 SSIYKKEGEVSEFSLAFLVEKLTSGGELIWMQERASSSSKSWITPDLKKEVSVKRYT 240
QY 268 QDPKIQWEEKPLPHLTLPQALPOVYGSNLTALAEKTKLHOENVLYVMAAQTQKSLT 327
DB 241 QDPKIQWEEKPLPHLTLPQALPHYAGSNLTALAEKTKLHOENVLYVMAAQTQKSLT 300

```

Qy	328	CEWNGPSPRLTMSTLKEKKAAYSRKRPVWVNLPEAGMOCILSSGVILESNIKVL	366
Db	301	CEWNGPSPRLTMSTLKEKKAAYSRKRPVWVNLPEAGMOCILSSGVILESNIKVL	366
Qy	388	PTWSTPVEP	396
Db	361	PTWPTPVPQ	369
RESULT 9			
ID	CD4_CERTO	STANDARD;	PRT; 397 AA.
AC	008336;		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	10-OCT-2003 (Rel. 42, Last annotation update)		
DE	T-cell surface glycoprotein CD4 (T-cell surface antigen T4/Leu-3) (Fragment).		
GN	CD4.		
OS	Cercopithecus torquatus alye (Red-crowned mangabey) (Sooty mangabey).		
OC	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;		
OC	Cercopitheciinae; Cercopithecus.		
OX	NCBI_TaxID=9531;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Blood.		
RX	MEDLINE=93049640; PubMed=1425921;		
RA	Fomsgaard A., Hirsch V.M., Johnson P.R.;		
RT	"Cloning and sequences of primate CD4 molecules: diversity of the		
RT	cellular receptor for simian immunodeficiency virus/human		
RT	immunodeficiency virus."		
RL	Eur. J. Immunol. 22:2973-2981(1992).		
CC	-1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell		
CC	receptor interaction. May regulate T-cell activation.		
CC	-1- SUBUNIT: Associates with p56-lck (By similarity).		
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.		
CC	-1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.		
CC	-1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	at the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; X73328; CAAS1754.1; -.		
DR	EMBL; X73327; CAAS1753.1; -.		
DR	HSPR; P01730; IM1Q.		
DR	GO; GO:0042101; C: T-cell receptor complex; ISS.		
DR	GO; GO:0015026; F: coreceptor activity; ISS.		
DR	GO; GO:0042289; F: MHC class II protein binding; ISS.		
DR	GO; GO:0006955; P: immune response; ISS.		
DR	GO; GO:0004508; P: positive regulation of interleukin-2 biosyn. . . ; ISS.		
DR	GO; GO:0030217; P: T-cell differentiation; ISS.		
DR	GO; GO:0045058; P: T-cell selection; ISS.		
DR	GO; GO:0007169; P: transmembrane receptor protein tyrosine kin. . . ; ISS.		
DR	InterPro; IPR000973; CD4_TCMG.		
DR	InterPro; IPR007110; Ig-like.		
DR	InterPro; IPR003596; Ig_V.		
DR	Pfam; PF00047; Ig_2.		
DR	PRINTS; PRO0692; CD4TCANTIGN.		
DR	SMART; SM00406; IGV; 1.		
DR	PROSITE; PSS0835; IG_LIKE; 1.		
KW	Immunoglobulin domain, Transmembrane, Glycoprotein, T-cell,		
KW	Immune response, Repeat, Lipoprotein, Palmitate.		
FT	NON TER	1	
FT	DOMAIN	<1	369
FT	TRANSMEM	370	391
FT	DOMAIN	392	>397
FT			EXTRACELLULAR (POTENTIAL).
FT			POTENTIAL.
FT			CYTOLASMIC (POTENTIAL).

```

FT DOMAIN <1 98 IG-LIKE V-TYPE.
FT DOMAIN 99 176 IG-LIKE C2-TYPE 1.
FT DOMAIN 177 290 IG-LIKE C2-TYPE 2.
FT DOMAIN 291 347 IG-LIKE C2-TYPE 3.
FT CARBOHYD 15 15 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 14 82 BY SIMILARITY.
FT DISULFID 128 157 BY SIMILARITY.
FT DISULFID 301 343 BY SIMILARITY.
FT LIPID 392 392 S-palmitoyl cysteine (By similarity).
FT LIPID 395 395 S-palmitoyl cysteine (By similarity).
FT LIPID 20 20 MISSING.
FT VARIANT 43 43 T -> I.
FT VARIANT 86 86 N -> D.
FT VARIANT 96 96 F -> L.
FT VARIANT 173 173 V -> M.
FT VARIANT 316 316 R -> K.
FT NON_TER 397 397
SQ SEQUENCE 397 AA; 43926 MW; 8660B636D2B38A7 CRC64;

```

```

Query Match 50.3%; Score 1716; DB 1; Length 397;
Best Local Similarity 89.4%; Pred. No. 6,3e-103;
Matches 330; Conservative 19; Mismatches 20; Indels 0; Gaps 0;

```

```

QY 28 VVLGGKGDVVELTCTASQKSIQFHMKNNSQIKILGNQSFLLKGPSKLANDRADSRRLW 87
DB 1 VVLGGKGDVVELACNASQKSKSQFWMKNSKQIKILGNQSFLLKGPSKLANDRADSRRLW 60
QY 88 DQGNFLLIKLIKIEDSDTYICEVNDQKEVQLVFGITANSDFHLLGOSLITLLESP 147
DB 61 DQGSMTIKIKLKEDSEYICEVENKKEVELLVFGITANSDFHLLGOSLITLLESP 120
QY 148 GSSPSVQCRSPKGNIOGKTLVSQLELQDSGTCTVLONOKVEFKIDIVLAFQKA 207
DB 121 GSSPSVKRSPKGNIOGKTLVSQLELQDSGTCTVLONOKVEFKIDIVLAFQKA 180
QY 208 SSIYKKEGEVSEFPLAFVTEKLTGSGELMWQERASSSKSWITFDLKNKEVSVKRV 267
DB 181 SSTYVKKGEVSEFPLAFVTEKLTGSGELMWQERASSSKSWITFDLKNKEVSVKRV 240
QY 268 QDPKIQMGKPLPLHITLPOALPOVAGSNTLTALEAKTGKLEHVNVLVWRATOLQKNT 327
DB 241 QDPKIQMGKPLPLHITLPOALPOVAGSNTLTALEAKTGKLEHVNVLVWRATOLQKNT 300
QY 328 CEVWGTPSPKMLSLKLENKAKYKREKPVVNLPEKGMOCCLSDSGVLLSNIKVYL 387
DB 301 CEVWGTPSPKMLSLKLENKAKYKREKPVVNLPEKGMOCCLSDSGVLLSNIKVYL 360
QY 388 PTWSTPEP 396
DB 361 PTWSTPEP 369

```

```

RESULT 10
CD4_SAISC STANDARD; PRT; 457 AA.
AC 029037:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3).
GN CD4.
OS Saimiri sciureus (Common squirrel monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Saimiri.
OX NCBI_Taxid=9521;
RN [1]
RP SEQUENCE FROM N.A.
RA Tatemura M., Hashimoto O.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
-i- FUNCTION: Accessory protein for MHC class-II antigen/T-cell

```

```

CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D86588; BAA13131.1; -.
CC HSSP; P01730; 1MBR.
DR GO; GO:0042101; C:T-cell receptor complex; ISS.
DR GO; GO:0015026; F:coreceptor activity; ISS.
DR GO; GO:0042289; F:MHC class II protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR009713; CD4_TcR.
DR InterPro; IPR007110; Ig_Like.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TcRANTIGEN.
DR SMART; SM00406; IgV; 1.
DR PROSITE; PS50835; IG_Like; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
KW SIGNAL

```

```

FT CHAIN 26 457 T-CELL SURFACE GLYCOPROTEIN CD4.
FT DOMAIN 26 395 EXTRACELLULAR (POTENTIAL).
FT TRANSEM 396 417 POTENTIAL.
FT DOMAIN 418 457 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 26 125 IG-LIKE V-TYPE.
FT DOMAIN 126 202 IG-LIKE C2-TYPE 1.
FT DOMAIN 203 316 IG-LIKE C2-TYPE 2.
FT DOMAIN 317 373 IG-LIKE C2-TYPE 3.
FT CARBOHYD 254 254 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 41 109 BY SIMILARITY.
FT DISULFID 155 184 BY SIMILARITY.
FT DISULFID 327 369 BY SIMILARITY.
FT LIPID 418 418 S-palmitoyl cysteine (By similarity).
FT LIPID 421 421 S-palmitoyl cysteine (By similarity).
SQ SEQUENCE 457 AA; 50871 MW; 57BED6344005A015 CRC64;

```

```

Query Match 46.3%; Score 1581.5; DB 1; Length 457;
Best Local Similarity 78.0%; Pred. No. 3,2e-94;
Matches 309; Conservative 32; Mismatches 54; Indels 1; Gaps 1;

```

```

QY 1 MNRGVPFRHLVLTQALPAATQGNKVVLGKGGDTVELCTASQKSIQFHMKNNSQIK 60
DB 1 MNRGVPFRHLVLTQALPAATQGNKVVLGKGGDTVELCTASQKSIQFHMKNNSQIK 60
QY 11LGNQSFLLKGPSKLANDRADSRRLWQGNFPLIKLIKIEDSDTYICEVNDQKEVQL 120
DB 61 ILGQNVFVTRGSGKLTDRIDSKSSMDRGSFPLLIKLIKIEDSEYICEVENKKEVEL 120
QY 121 LVFGITANSDFHLLGOSLITLLESPGSSPVQCRSPKGNIOGKTLVSQLELQDSG 180
DB 121 QVFGITANSDFHLLGOSLITLLESPGSSPVQCRSPKGNIOGKTLVSQLELQDSG 180
QY 181 TWTCVTVQNGKXVEFKIDIVLAFQKASSVYKKEGEVSEFPLAFVTEKLTGSGELMW 240
DB 181 TWTCVTVQNGKXVEFKIDIVLAFQKASSVYKKEGEVSEFPLAFVTEKLTGSGELMW 239
QY 241 QABRASSSKSWITFDLKNKEVSVKRVTPDKLQMGKPLPLHITLPOALPOVAGSNTLTA 300

```

|||||
Db 240 QAEKSSSSKSWITFNLTKEQVYKVLVTQPKLRMGKGLPLHLTLAQLPQVAGSGNFTLA 299
301 LEAKTKGLQEOVNLVYMRATQLOKNTCTCEWGPTSKMLSLKLEKKEKSKREKPVWV 360
300 LKGTGKGLQEOVNLVYMRATQLOKNTCEWGPTSKMLSLKLEKKEKSKREKAVWV 359
Qy 361 LNPEAGMOCCLSDSGVLLSENIKVLPTWSTPEV 396
Db 360 LNPEAGMOCCLSDSGVLLSENIKVLPTWSTPEV 395
RESULT 11
GCL_HUMAN STANDARD; PRT; 330 AA.
AC POL557;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE IG gamma-1 chain C region.
GN IGHG1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82274238; PubMed=6287432;
RA Ellison J.W., Berson B.J., Hood L.E.;
RT "The nucleotide sequence of a human immunoglobulin C gamma1 gene.";
RL Nucleic Acids Res. 10:4071-4079(1982).
RN [2]
RP SEQUENCE OF 1-135 (MYELOMA PROTEIN EU).
RX MEDLINE=71064024; PubMed=5489771;
RA Cunningham B.A., Rutishauser U., Gall W.E., Gottlieb P.D.,
RA Waxedal M.J., Edelman G.M.;
RT "The covalent structure of a human gamma G-immunoglobulin. VII. Amino
RT acid sequence of heavy-chain cyanogen bromide fragments H1-H4.";
RL Biochemistry 9:3161-3170(1970).
RN [3]
RP SEQUENCE OF 136-329 (EU).
RX MEDLINE=71064025; PubMed=5530842;
RA Rutishauser U., Cunningham B.A., Bennett C., Konigsberg W.H.,
RA Edelman G.M.;
RT "The covalent structure of a human gamma G-immunoglobulin. 8. Amino
RT acid sequence of heavy-chain cyanogen bromide fragments H5-H7.";
RL Biochemistry 9:3171-3181(1970).
RN [4]
RP SEQUENCE (MYELOMA PROTEIN NIE).
RX MEDLINE=77070269; PubMed=826475;
RA Ponstingl H., Hilschmann N.;
RT "The rule of antibody structure. The primary structure of a
RT monoclonal IgG1 immunoglobulin (myeloma protein Nie). III. The
RT chymotryptic peptides of the H-chain, alignment of the tryptic
RT peptides and discussion of the complete structure.";
RL Hoppe-Seyler's Z. Physiol. Chem. 357:1571-1604(1976).
RN [5]
RP SEQUENCE (MYELOMA PROTEIN KOL), AND DISULFIDE BONDS.
RX MEDLINE=83289131; PubMed=6884994;
RA Schmidt W.E., Jung H.-D., Palm W., Hilschmann N.;
RT "Three-dimensional structure determination of antibodies. Primary
RT structure of crystallized monoclonal immunoglobulin IgG1 KOL, I.";
RL Hoppe-Seyler's Z. Physiol. Chem. 364:713-747(1983).
RN [6]
RP DISULFIDE BONDS.
RX MEDLINE=71064027; PubMed=4923144;
RA Gall W.E., Edelman G.M.;
RT "The covalent structure of a human gamma G-immunoglobulin. X.
RT interchain disulfide bonds.";
RL Biochemistry 9:3188-3196(1970).
RN [7]
RP DISULFIDE BONDS.
RX MEDLINE=77070267; PubMed=1002129;

RA Dreker L., Schwarz J., Reichel W., Hilschmann N.;
RT "Rule of antibody structure. The primary structure of a monoclonal
RT IgG1 immunoglobulin (myeloma protein Nie). I: Purification and
RT characterization of the protein, the L- and H-chains, the
RT cyanogen bromide cleavage products, and the disulfide bridges.";
RL Hoppe-Seyler's Z. Physiol. Chem. 357:1515-1540(1976).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).
RX MEDLINE=81208100; PubMed=7236608;
RA Deisenhofer J.;
RT "Crystallographic refinement and atomic models of a human Fc fragment
RT and its complex with fragment B of protein A from Staphylococcus
RT aureus at 2.9- and 2.8-A resolution.";
RL Biochemistry 20:2361-2370(1981).
CC -I- MISCELLANEOUS: Nie has the G1M(17) allotypic marker, 97-K, and the
CC G1M(1) markers, 239-D and 241-L. KOL and EU sequences have the
CC G1M(3) marker and the G1M (non-1) markers.
CC -I- MISCELLANEOUS: Nie also differs in the amidation states of
CC 35, 116, 198, 269 and 272.
CC -I- MISCELLANEOUS: EU also differs in the amidation states of residues
CC 155, 166, 177, 195, 198, 269, and 272 and in the order of residues
CC 268-272.
CC -I- MISCELLANEOUS: KOL also differs in the amidation states of
CC residues 198, 267 and 272.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib.ch).
CC -----
CC EMBL: J00228; AAC82527.1; ALT_INT.
CC PIR: A93433; GHU.
DR PDB: 1FC1; 15-JUL-92.
DR PDB: 1FC2; 15-JUL-92.
DR PDB: 1A77; 12-NOV-97.
DR PDB: 1D5B; 09-FEB-00.
DR PDB: 1D5T; 09-FEB-00.
DR PDB: 1D6V; 04-OCT-00.
DR PDB: 1DN2; 17-MAY-00.
DR PDB: 1E4K; 06-JUN-01.
DR PDB: 1FCC; 20-JUL-95.
DR PDB: 1H2H; 12-JUN-02.
DR PDB: 1I7Z; 08-AUG-01.
DR PDB: 1IIS; 16-MAY-01.
DR PDB: 1IIX; 16-MAY-01.
DR PDB: 1L6X; 10-APR-02.
DR PDB: 2RCS; 12-NOV-97.
DR GeneW; HGNC:5525; IGHG1.
DR MIM; 147100; -.
DR GO; GO:0005624; C:membrane fraction; NAS.
DR GO; GO:0003823; F:antigen binding; TAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig-cl.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_3.
DR SMART; SM00407; IGc1; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 2.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW 3D-structure.
FT NON_TER 1 1
FT DOMAIN 1 98 CH1.
FT DOMAIN 99 110 HINGE.
FT DOMAIN 111 223 CH2.
FT DOMAIN 224 330 CH3.
FT DISULFID 27 83
FT DISULFID 103 103
FT DISULFID 109 109 INTERCHAIN (WITH LIGHT CHAIN).
FT DISULFID 112 112 INTERCHAIN (WITH HEAVY CHAIN).

```

FT DISULFID 144 204
FT DISULFID 250 308
FT CARBOHYD 180 180
FT MOD RES 330 330
FT VARIANT 97 97
FT VARIANT 239 239
FT VARIANT 241 241
FT VARIANT 241 241
FT STRAND 122 126
FT HELIX 136 134
FT TURN 136 137
FT STRAND 141 147
FT STRAND 157 162
FT TURN 163 164
FT STRAND 165 166
FT TURN 168 171
FT STRAND 176 179
FT TURN 180 181
FT STRAND 182 190
FT HELIX 193 197
FT TURN 198 199
FT STRAND 202 207
FT TURN 209 210
FT STRAND 215 219
FT STRAND 227 227
FT STRAND 230 234
FT HELIX 238 242
FT STRAND 245 256
FT STRAND 260 265
FT STRAND 270 270
FT STRAND 274 276
FT STRAND 280 281
FT TURN 283 284
FT STRAND 287 296
FT HELIX 297 301
FT TURN 302 303
FT STRAND 305 312
FT TURN 313 314
FT TURN 316 317
FT STRAND 320 325
SQ SEQUENCE 330 AA; 36106 MW; 3770EB106C2FA3D CRC64;

Query Match 37.1%; Score 1266; DB 1; Length 330;
Best Local Similarity 96.7%; Pred. No. 3.5e-74;
Matches 236; Conservative 0; Mismatches 4; Indels 4; Gaps 1;

```

```

DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE IG gamma-3 chain C region (Heavy chain disease protein) (HDC).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE (DISEASE PROTEIN WIS).
RX MEDLINE=81021546; PubMed=6774747;
RA Frangione B., Rosenwasser E., Prelli F., Franklin E.C.;
RT "Primary structure of human gamma 3 immunoglobulin deletion mutant:
RL gamma 3 heavy-chain disease protein WIS."
RN Biochemistry 19:4304-4308(1980).
RP REVISIONS TO 12-97 (PROTEIN WIS).
RX MEDLINE=77118561; PubMed=402363;
RA Michaelson T.E., Frangione B., Franklin E.C.;
RT "Primary structure of the 'hinge' region of human IgG3. Probable
RL quadruplication of a 15-amino acid residue basic unit."
RN J. Biol. Chem. 252:883-889(1977).
RP REVISIONS TO 59-289 (PROTEIN WIS/DISEASE PROTEIN ZUC).
RX MEDLINE=77021516; PubMed=823945;
RA Wolfenstein-Todel C., Frangione B., Prelli F., Franklin E.C.;
RT "The amino acid sequence of 'heavy chain disease' protein ZUC.
RL Structure of the Fc fragment of immunoglobulin G3."
RN Biochem. Biophys. Res. Commun. 71:907-914(1976).
RP SEQUENCE FROM N.A. (DISEASE PROTEIN OMM).
RX MEDLINE=82247835; PubMed=6808505;
RA Alexander A., Steinmetz M., Barilault D., Frangione B.,
RA Franklin E.C., Hood L., Buxbaum J.N.;
RT "Gamma heavy chain disease in man: cDNA sequence supports partial
RL gene deletion model."
RN Proc. Natl. Acad. Sci. U.S.A. 79:3260-3264(1982).
CC -1- SUBUNIT: Dimer linked by 12 disulfide bonds; it has an extra
interchain disulfide bond at position 7 in addition to the 11
normally present in the hinge region.
CC -1- MISCELLANEOUS: The heavy chain disease protein WIS is shown.
CC -1- MISCELLANEOUS: The sequence of residues 42-76 was taken from the
Ref.2.
CC -1- MISCELLANEOUS: Disease protein WIS is lacking most of the V region
and all of the CH1 region.
CC -1- MISCELLANEOUS: Disease protein ZUC lack most of the V region, all
of the CH1 region, and part of the hinge compared with normal
gamma-3 heavy chains.
CC -1- MISCELLANEOUS: Disease protein OMM may represent an allelic form
or another gamma chain subclass.
CC -1- MISCELLANEOUS: The hinge region in gamma-3 chains is about four
times as long as in other gamma chains and contains three
identical 15-residue segments preceded by a similar 17-residue
segment (12-28).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL, outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J00231; AA52805.1; ALT_SEQ.
DR HSSP; P01857; IFC1.
DR Genew; HGNC:5527; IGNG3.
DR MIM; 147120; -.
DR GO; GO:0005624; C:membrane fraction; NMS.
DR GO; GO:0003823; P:antigen binding; TAS.
DR GO; GO:0006955; P:immune response; NMS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_c1.

```

```

DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig1; 2.
DR SMART: SM00407; Igcl; 1.
DR PROSITE: PSS0835; IG_LIKE; 2.
DR PROSITE: PSS0290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat;
KW Pyridoxal carboxylic acid.
FT DOMAIN 12 73 HINGE.
FT DOMAIN 74 183 CH2.
FT DOMAIN 184 289 CH3.
FT REPEAT 29 43
FT REPEAT 44 58
FT REPEAT 59 73
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT CARBOHYD 6 6 N-LINKED (GLCNAC...).
FT DISULFID 7 7 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT DISULFID 24 24 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT DISULFID 27 27 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT DISULFID 33 33 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT DISULFID 39 39 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT DISULFID 42 42 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT DISULFID 48 48 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT DISULFID 54 54 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT DISULFID 57 57 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT DISULFID 63 63 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT DISULFID 69 69 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT DISULFID 72 72 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT CARBOHYD 140 140 N-LINKED (GLCNAC...).
FT MOD_RES 290 290 REMOVED POST-TRANSLATIONALLY.
FT VARIANT 126 127 QV -> EB (IN ZUC).
FT VARIANT 134 134 /FTID=VAR 003890.
FT VARIANT 139 139 P -> L (IN OMM).
FT VARIANT 182 182 F -> Y (IN OMM).
FT VARIANT 227 227 /FTID=VAR 003892.
FT VARIANT 227 227 T -> A (IN OMM).
FT VARIANT 227 227 /FTID=VAR 003893.
FT VARIANT 227 227 S -> N (IN OMM).
FT VARIANT 227 227 /FTID=VAR 003894.
FT VARIANT 227 227 MISSING (IN ZUC).
FT VARIANT 279 279 /FTID=VAR 003895.
FT VARIANT 279 279 F -> Y (IN OMM).
FT VARIANT 279 279 /FTID=VAR 003896.
SQ SEQUENCE 290 AA; 32331 MW; B69CB35705B2F46 CRC64;

Query Match 33.7%; Score 1151; DB 1; Length 290;
Best Local Similarity 90.5%; Pred. No. 6.9e-67;
Matches 209; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

QY 395 EPKSCDHTHTCPGPCAPPELLGSPVFLFPKPKDITLMISRTPEVTCVVDVSHDEPKV 454
DB 59 EPKSCDTPPCPCPCAPPELLGSPVFLFPKPKDITLMISRTPEVTCVVDVSHDEPKV 118
QY 455 NMVYDGEVHNNAKTKRREQYNSTYVSVLTVLADQWLNKGEYKCKVSNKLPAPRIEKT 514
DB 119 KMVYDGVQVHNNAKTKRREQYNSTYVSVLTVLADQWLNKGEYKCKVSNKLPAPRIEKT 178
QY 515 ISAKAQPRBPQYITTPPSRDELTKQVSLTCLVKGFYSDDIAVENESGQPPNNYKTP 574
DB 179 ISITKQPRBPQYITTPPSRDEETKQVSLTCLVKGFYSDDIAVENESGQPPNNYKTP 238
QY 575 PVLDSGSEFLYSKLTVDKSRMQQGVFSCSVNHEALNHNHYOKSLSPG 625
DB 239 PVLDSGSEFLYSKLTVDKSRMQQGVFSCSVNHEALNHNHYOKSLSPG 289

RESULT 13
GC2 HUMAN STANDARD; PRT; 326 AA.
AC P01859;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE Ig gamma-2 chain C region.
GN IGH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE OF 2-326 FROM N.A.
RX MEDLINE=82197621; PubMed=6804948;
RA Elison J.W., Hood L.E.;
RT "Linkage and sequence homology of two human immunoglobulin gamma
RT heavy chain constant region genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:1984-1988(1982).
RN [2]
RP SEQUENCE OF 88-115 FROM N.A.
RC TISSUE=Petal liver;
RX MEDLINE=83001943; PubMed=6811139;
RA Takahashi N., Ueda S., Obata M., Nikaido T., Nakai S., Honjo T.;
RT "Structure of human immunoglobulin gamma genes: implications for
RT evolution of a gene family.";
RL Cell 29:671-679(1982).
RN [3]
RP SEQUENCE OF 99-177 AND 310-326 FROM N.A.
RC TISSUE=Petal liver;
RX MEDLINE=84235992; PubMed=6329676;
RA Krawinkel U., Rabbits T.H.;
RT "Comparison of the hinge-coding segments in human immunoglobulin gamma
RT heavy chain genes and the linkage of the gamma 2 and gamma 4 subclases
RT genes.";
RL EMBO J. 1:403-407(1982).
RN [4]
RP SEQUENCE OF 1-325 (MYELOMA PROTEIN TIL).
RX MEDLINE=81007873; PubMed=6774012;
RA Wang A.-C., Tung E., Fudenberg H.H.;
RT "The primary structure of a human IgG2 heavy chain: genetic,
RT evolutionary, and functional implications.";
RL J. Immunol. 125:1048-1054(1980).
RN [5]
RP SEQUENCE OF 1-85 AND 133-325 (MYELOMA PROTEIN ZIE).
RX MEDLINE=80001357; PubMed=113060;
RA Connell G.E., Parr D.M., Hofmann T.;
RT "The amino acid sequences of the three heavy chain constant region
RT domains of a human IgG2 myeloma protein.";
RL Can. J. Biochem. 57:758-767(1979).
RN [6]
RP SEQUENCE OF 238-275 (ZIE).
RX MEDLINE=80114419; PubMed=118920;
RA Hofmann T., Parr D.M.;
RT "A note of the amino acid sequence of residues 381-391 of human
RT immunoglobulin gamma chains";
RL Mol. Immunol. 16:923-925(1979).
RN [7]
RP REVISIONS TO 25; 59; 60 AND 264-268 (ZIE).
RA Hofmann T., Parr D.M.;
RT Submitted (MAR-1980) to the PIR data bank.
RN [8]
RP SEQUENCE OF 1-121 (DOT).
RX MEDLINE=95255298; PubMed=7737190;
RA Stoppin M., Bellotti V., Negri A., Merlini G., Garver F., Ferri G.;
RT "Characterization of the two unique human anti-flavin monoclonal
RT immunoglobulins";
RL Eur. J. Biochem. 228:886-893(1995).
RN [9]
RP DISULFIDE BONDS.
RX MEDLINE=72033500; PubMed=4940472;
RA Milestein C., Frangione B.;
RT "Disulphide bridges of the heavy chain of human immunoglobulin G2.";
RL Biochem. J. 121:217-225(1971).
RN [10]
RP DISULFIDE BONDS.
RX MEDLINE=69064124; PubMed=5782707;
RA Frangione B., Milestein C., Pink J.R.L.;
RT "Structural studies of immunoglobulin G2.";

```

```
RL Nature 221:145-148 (1969).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: J00230; AAB59393.1; -.
DR PIR: A93906; G2HU.
DR HSSP: P01857; 1FCL.
DR Genew: HGNC:5526; IGHG2.
DR MIM: 147110; -.
DR GO: GO:0005624; C:membrane fraction; NAS.
DR GO: GO:0003823; F:antigen binding; TAS.
DR GO: GO:0006955; P:immune response; NAS.
DR InterPro: IPR007110; IG-like.
DR InterPro: IPR003597; IG_c1.
DR InterPro: IPR003006; IG_MHC.
DR Pfam: PF00047; Ig; 3.
DR SMART: SM00407; IGc1; 2.
DR PROSITE: PS00835; IG_LIKE; 3.
DR PROSITE: PS00290; IG_MHC; 2.
KW Immunoglobulin domain; Immunoglobulin C region.
FT DOMAIN 1 98 CH1.
FT 99 110 HINGE.
FT 111 219 HINGE.
FT DOMAIN 220 326 CH3.
FT DISULFID 14 14 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 27 83 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 102 102 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 103 103 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 140 200
FT DISULFID 246 304
FT DISULFID 156 156
FT SITE 326 326
FT MOD RES 60 60
FT VARIANT 60 60
FT CONFLICT 109 109
FT SEQUENCE 326 AA; 35884 MW; 83108786C68CF9C CRC64;
Query Match 33.7%; Score 1150; DB 1; Length 326;
Best Local Similarity 88.9%; Pred. No. 9,3e-67;
Matches 217; Conservative 9; Mismatches 10; Indels 8; Gaps 3;
QY 382 SNIKYLPMTSTPVEKSCDKTHTCPCPAPPELLGGSVFLFPPKPKDTLMISRPPEVTCV 441
DB 90 SNTKVDKTI-----VERKCKVE---CPKCAPP-VAGPSVFLFPPKPKDTLMISRPPEVTCV 141
QY 442 VVDVSHEDPEVKEFNMYVDGVEVHNATKYPREQYNSTYRVASVLTVLHQMVLNGKEYCK 501
DB 142 VVDVSHEDPEVQFNMYVDGVEVHNATKYPREQYNSTYRVASVLTVLHQMVLNGKEYCK 201
QY 502 VSNKALPAPIKKTISKAGQPREPOVYTLPPSRDLTKNOVSLTCLVNGFPPSDIAVWE 561
DB 202 VSNKGLPAPIKKTISKAGQPREPOVYTLPPSRDLTKNOVSLTCLVNGFPPSDIAVWE 261
QY 562 SNGQENNNYKTTTPYLDSDGSFFLYSKLTVDSKRMQGNVSCSVMEALNNHTQKSL 621
DB 262 SNGQENNNYKTTTPYLDSDGSFFLYSKLTVDSKRMQGNVSCSVMEALNNHTQKSL 321
QY 622 LSPG 625
DB 322 LSPG 325
```

```
ID CD4_RABIT STANDARD; PRT; 459 AA.
AC P46630;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3).
GN CD4.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OC NCBI_TaxID=9986;
RX MEDLINE=92390370; PubMed=1518821;
RA Hague B.F., Sawasdi Kosol S., Brown T.J., Lee K., Recker D.P.,
RA Kindt T.J.;
RT "CD4 and its role in infection of rabbit cell lines by human
RT immunodeficiency virus type 1."
RL Proc. Natl. Acad. Sci. U.S.A. 89:7963-7967(1992).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (by similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: M92840; AAA31198.1; -.
DR PIR: A46254; A46254.
DR HSSP: P01730; 1MBR.
DR GO: GO:0042101; C:T-cell receptor complex; ISS.
DR GO: GO:0015026; F:coreceptor activity; ISS.
DR GO: GO:0042289; F:MHC class II protein binding; ISS.
DR GO: GO:0006955; P:immune response; ISS.
DR GO: GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO: GO:0030217; P:T-cell differentiation; ISS.
DR GO: GO:0045058; P:T-cell selection; ISS.
DR GO: GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro: IPR000973; CD4_TcRg.
DR InterPro: IPR007110; IG_LIKE.
DR InterPro: IPR003596; IG_V.
DR Pfam: PF00047; Ig; 2.
DR PRINTS: PR00692; CD4TCANTIGEN.
DR SMART: SM00406; IG; 1.
DR PROSITE: PS00835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 459
FT DOMAIN 26 396
FT TRANSLEM 397 419
FT DOMAIN 420 459
FT DOMAIN 26 129
FT DOMAIN 130 308
FT DOMAIN 209 318
FT DOMAIN 319 374
FT CARBOHYD 299 299
FT DISULFID 41 113
FT DISULFID 329 370
FT DISULFID 420 420
FT LIPID 423 423
FT SEQUENCE 459 AA; 50886 MW; B323311CDBD40013D CRC64;
Query Match 33.7%; Score 1149; DB 1; Length 459;
Best Local Similarity 56.9%; Pred. No. 1.7e-66;
```



```

RP SEQUENCE FROM N.A.
RC STRAIN=Beagle; TISSUE=Thymus;
RX MEDLINE=94378217; PubMed=8091416;
RA Gorman S.D., Frewin M.R., Cobbold S.P., Waldmann H.;
RT "Isolation and expression of cDNA encoding the canine CD4 and CD8
  alpha antigens."
RL Tissue Antigens 43:184-188(1994).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
  receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: Expressed in macrophages and a subset of
  T lymphocytes.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L06130; AAB02295.1; -.
DR EMBL; X68565; -. NOT_ANNOTATED_CDS.
DR HSSP; P01730; 1MER.
DR GO; GO:0042101; C:T-cell receptor complex; ISS.
DR GO; GO:0015026; F:coreceptor activity; ISS.
DR GO; GO:0042289; F:MHC class II protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. .; ISS.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig-TcAg.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 3.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
  Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 24
FT CHAIN 25 463 T-CELL SURFACE GLYCOPROTEIN CD4.
FT DOMAIN 25 401 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 402 423 POTENTIAL.
FT DOMAIN 424 463 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 26 124 IG-LIKE V-TYPE.
FT DOMAIN 125 211 IG-LIKE C2-TYPE 1.
FT DOMAIN 212 321 IG-LIKE C2-TYPE 2.
FT DOMAIN 322 378 IG-LIKE C2-TYPE 3.
FT DISULFID 41 109 BY SIMILARITY.
FT DISULFID 332 374 BY SIMILARITY.
FT LIPID 424 424 S-palmitoyl cysteine (By similarity).
FT LIPID 427 427 S-palmitoyl cysteine (By similarity).
FT CARBOHYD 123 123 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 176 176 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 389 389 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 463 AA; 51639 MW; 95805170CB44A833 CRC64;
Query Match 33.4%; Score 1141; DB 1; Length 463;
Best Local Similarity 57.6%; Pred. No. 5,6e-66;
Matches 236; Conservative 63; Mismatches 93; Indels 18; Gaps 6;
QY 1 MNRGVPFPHLLVLTALPAATGKNGKVLGKGGDTVLTCTAGSKSGIOFHWNKSNQIK 60
DB 1 MNOGAFAFHLILMLQVLNLPATVPVREVLVKGAGAVELPCOTSQKKNHFNWDDSSVQO 60

```

```

QY 61 ILNGQSGFLTKGPKLNDRADSRRLMDQGNFLLINKLKIEDSDTYICEVEDQKEEVQL 120
DB 61 ILNGQSGFWTVGSSRLRHGRVSKSKNMLMDQGSFPLVIDLEAVDSGIFYCOT-DRQGEVLE 119
QY 121 LVFGLTL-----NSDTHLLOGSLTITLSEPPSSPSVOCRSRGKNIOGKTLNV 171
DB 120 LVFNLTKMKDSSGSSSSNRLRLOGQOLTITLLEPSSSPSVQKGGNNSKKGQNLSTL 179
QY 172 SQLELDGSGTWCTCVLONOKKVEFKIDIVLAFOKASSIYVKKGEOVFSPFAFYVEK 231
DB 180 SWEPLQGGTWTCTLISSQKTEVEPNIVLAIPKVSNTYFARGDQVRSFPLSFEDEN 239
QY 232 LYSGSELMOAERASSSSKSWITFDLKNKEVSVKRVTPDKLQWKKPLHLTLPOALPYQ 291
DB 240 LV-GEIRWQAQAGASSSLWISFTLENRKLSMKEAHAFLKIQMKESLPLRFTLPQVLSRY 297
QY 292 AGSNLTLLAEKATKGLHGVNLYVMAATOLQKLTETWGPSPKMLSLKENKAKV 351
DB 298 AGSGILTLNL-AK-GTLVQEVNLYVMRANSSQNNLTCEVIGPSPSLTSLNLEQAAKV 355
QY 352 SKREKPYVWNPBAGWMOCLLSDSGVLYLESNIKVLPTWSPVPEKSCDK 401
DB 356 SKQKLVWVDPBEGTWQCLLSDKDKVLLASSLVN----SSPVVYKSWPK 401
RESULT 17
CD4_RAT ID CD4_RAT STANDARD; PRT; 457 AA.
AC P05540;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
  T4/Leu-3) (W3/25 antigen).
GN CD4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=8717535; PubMed=3104900;
RA Clark S.D., Jeffries W.A., Barclay A.N., Gagnon J., Williams A.F.;
RT "Peptide and nucleotide sequences of rat CD4 (W3/25) antigen:
  evidence for derivation from a structure with four
  immunoglobulin-related domains."
RL Proc. Natl. Acad. Sci. U.S.A. 84:1649-1653(1987).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 210-393.
RX MEDLINE=93262437; PubMed=8493535;
RA Brady R.L., Dodson E.J., Dodson G.G., Lange G., Davis S.J.,
  Williams A.F., Barclay A.N.;
RT "Crystal structure of domains 3 and 4 of rat CD4: relation to the
  NH2-terminal domains."
RL Science 260:979-983(1993).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
  receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M15768; AAA40901.1; -.
DR PIR; A27449; A27449.
DR PDB; 1CID; 15-JUL-93.

```



```
DR GlycoSiteDB; P05540; C: T-cell receptor complex; ISS.
DR GO: GO:0042101; C: T-cell receptor activity; ISS.
DR GO: GO:0015026; F: coreceptor activity; ISS.
DR GO: GO:0042289; F: MHC class II protein binding; ISS.
DR GO: GO:0006955; P: immune response; ISS.
DR GO: GO:0045086; P: positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO: GO:0030217; P: T-cell differentiation; ISS.
DR GO: GO:0045058; P: T-cell selection; ISS.
DR GO: GO:0007169; P: transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro: IPR000973; CD4_TcRg.
DR InterPro: IPR007110; Ig-ILike.
DR InterPro: IPR003599; Ig.
DR Pfam: PF00047; Ig_2.
DR SMART: SM00409; Ig_2.
DR PROSITE: PS00835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate; 3D-structure.
FT CHAIN 1 28 457 T-CELL SURFACE GLYCOPROTEIN CD4.
FT SIGNAL 27 27 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 28 394 POTENTIAL.
FT TRANSMEM 395 417 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 418 457 POTENTIAL.
FT DOMAIN 128 127 IG-LIKE V-TYPE.
FT DOMAIN 128 206 IG-LIKE C2-TYPE 1.
FT DOMAIN 207 316 IG-LIKE C2-TYPE 2.
FT DOMAIN 317 374 IG-LIKE C2-TYPE 3.
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 43 111 BY SIMILARITY.
FT DISULFID 158 187 BY SIMILARITY.
FT DISULFID 328 370 BY SIMILARITY.
FT LIPID 418 418 S-palmitoyl cysteine (By similarity).
FT LIPID 421 421 S-palmitoyl cysteine (By similarity).
FT STRAND 213 217
FT STRAND 218 219
FT STRAND 222 225
FT STRAND 235 243
FT STRAND 252 258
FT STRAND 259 260
FT STRAND 261 265
FT STRAND 274 275
FT STRAND 278 278
FT TURN 279 280
FT STRAND 282 285
FT STRAND 290 292
FT HELIX 294 301
FT STRAND 306 319
FT STRAND 325 331
FT STRAND 338 344
FT TURN 345 346
FT STRAND 349 353
FT STRAND 357 361
FT STRAND 367 374
FT TURN 375 376
FT STRAND 377 385
SQ SEQUENCE 457 AA; 51437 MW; 477BE157D30954C1 CRC64;

Query Match 29.3%; Score 999; DB 1; Length 457;
Best Local Similarity 52.9%; Pred. No. 6,9e-57;
Matches 207; Conservative 61; Mismatches 115; Indels 8; Gaps 5;
```

```
Db 121 ELWFRVTNPGTRLLQGQSLLTLIDSNPKVSDPPECKHKSSNIVKDSKAFSTSLRIQ 180
Qy 178 DSGTTCYTLQOKKVEFKIDIVLAFOKASSIVYKKEGOVEFEPFLAFYKLTGSGE 237
Db 181 DSGINCTVTLNOKKGSFPMKLSVIGFASSTSTAVKSEBSAEFSPPLNIGESL--QGE 238
Qy 238 LMKQERASSSKSWITFDLKNKEVSKRVTOPDKLQMGKKLPLHLTLPOALPOVYSGNL 297
Db 239 LMKKEKAPSSQSWITFSLKNOKVOKSTSNPKQLSETLPLTLQIPOVSLQFAGSGL 298
Qy 298 TLAEAKTGKLRQEVNLVVRATOLQKN-LTCEVWGPTSPKMLSLKLENKAKYKREK 356
Db 299 TLTLD--RGIIVQEVNLVVRKVTQPSNTLTCEVWGPTSPKMRLLIKQENQEARVSRQEK 356
Qy 357 PYVYANPEPGMVOCLISDGOVLLESNIKVL 387
Db 357 VIQVAPENAGVQCILSBEGBEVMDSKIQVL 387

RESULT 18
CD4_MOUSE STANDARD; PRT; 457 AA.
ID CD4_MOUSE
AC P06332;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
T4/Leu-3) (T-cell differentiation antigen L3T4).
GN CD4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87018845; PubMed=3094146;
RA Tourville B., Gorman S.D., Field E.H., Hunkapiller T., Parnes J.R.;
RT "Isolation and sequence of L3T4 complementary DNA clones: expression
in T cells and brain.";
RL Science 234:610-614(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87115821; PubMed=3027575;
RA Littman D.R., Gettner S.N.;
RT "Unusual insertion in the immunoglobulin domain of the newly isolated
murine CD4 (L3T4) gene.";
RL Nature 325:453-455(1987).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=88152875; PubMed=3326818;
RA Parnes J.R., Hunkapiller T.;
RT "L3T4 and the immunoglobulin gene superfamily: new relationships
between the immune system and the nervous system.";
RL Immunol. Rev. 100:109-127(1987).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=88041159; PubMed=2823269;
RA Gorman S.D., Tourville B., Parnes J.R.;
RT "Structure of the mouse gene encoding CD4 and an unusual transcript
in brain.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:7644-7648(1987).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=98112780; PubMed=9445485;
RA Anasetti Lari M.A., Oeltjen J.C., Schwartz S., Zhang Z., Muzny D.M.,
Lu J.J., Correll J.H., Chinnault A.C., Belmont J.W., Miller W.,
Gibbs R.A.;
RT "Comparative sequence analysis of a gene-rich cluster at human
RT chromosome 12p13 and its syntenic region in mouse chromosome 6.";
RL Genome Res. 8:29-40(1998).
RN [6]
```

RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strusberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carinci P., Prange C.,
 RA Rabe S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Raeb S.A., McEwan P.J., McKernan K.J., Malik J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butcherfield J.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 [7]
 RP SEQUENCE OF 27-43.
 RX MEDLINE=86166694; PubMed=3082751;
 RX Classon B.J., Tsagaratos J., Kitzbaum L., Maddox J., McKay C.R.,
 RA Braden M., McKenzie I.F.C., Walker I.D.;
 RT "The L3T4 antigen in mouse and the sheep equivalent are
 RT immunoglobulin-like".
 RL Immunogenetics 23:129-132 (1986).
 [8]
 RP DISULFIDE BONDS.
 RX MEDLINE=86233454; PubMed=3086886;
 RX Classon B.J., Tsagaratos J., McKenzie I.F.C., Walker I.D.;
 RT "Partial primary structure of the T4 antigens of mouse and sheep:
 RT assignment of intrachain disulfide bonds.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:4499-4503 (1986).
 CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
 CC receptor interaction. May regulate T-cell activation.
 CC -1- SUBUNIT: Associates with p56-lick (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: Type I membrane protein.
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=P06332-1; Sequence=Displayed;
 CC Name=2; Synonyms=Brain-specific;
 CC IsoId=P06332-2; Sequence=VSP_002489;
 CC -1- SIMILARITY: Contains 3 Immunoglobulin-like C2-type domains.
 CC -1- SIMILARITY: Contains 1 Immunoglobulin-like V-type domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M36850; AAA39401.1; -;
 DR EMBL; M13816; AAA37267.1; -;
 DR EMBL; X04836; CAA28539.1; -;
 DR EMBL; M36851; AAA39402.1; -;
 DR EMBL; M17080; AAA37403.1; -;
 DR EMBL; M17078; AAA37403.1; JOINED.
 DR EMBL; M17079; AAA37403.1; JOINED.
 DR EMBL; AC002397; AAC36010.1; -;
 DR EMBL; BC039137; AAC39137.1; -;
 DR PIR; A02110; RWMST4.
 DR HSSP; P01730; 1WBR.
 DR MGD; MGI:88335; Cd4.
 DR GO; GO:0042101; C:T-cell receptor complex; ISS.
 DR GO; GO:0015026; F:coreceptor activity; ISS.

DR GO; GO:0042289; F:MHC class II protein binding; ISS.
 DR GO; GO:0006955; P:immune response; ISS.
 DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
 DR GO; GO:0030217; P:T-cell differentiation; ISS.
 DR GO; GO:0045058; P:T-cell selection; ISS.
 DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
 DR InterPro; IPR009733; CD4_TcRg.
 DR InterPro; IPR007110; Ig_V.
 DR InterPro; IPR003596; Ig_V.
 DR Pfam; PF00047; Ig_2.
 DR PRINTS; PR00692; CD4TCANTIGEN.
 DR SMART; SMO0406; IGV_1.
 DR PROSITE; PS50835; IG_LIKE_1.
 KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
 KW Immune response; Repeat; Signal; Lipoprotein; Palmitate;
 KW Alternative splicing.
 FT SIGNAL 1 26
 FT CHAIN 27 457
 FT DOMAIN 27 394
 FT TRANSMEM 395 417
 FT DOMAIN 418 457
 FT DOMAIN 27 128
 FT DOMAIN 129 207
 FT DOMAIN 208 317
 FT DOMAIN 318 374
 FT CARBOHYD 187 187
 FT CARBOHYD 298 298
 FT CARBOHYD 323 323
 FT CARBOHYD 392 392
 FT DISULFID 42 112
 FT DISULFID 159 188
 FT DISULFID 328 370
 FT LIPID 418 418
 FT LIPID 421 421
 FT VARSPLIC 1 240
 FT SEQUENCE 457 AA; 51296 MW; 1B1DA7527CB00F33 CRC64;
 Query Match 29.1%; Score 993; DB 1; Length 457;
 Best Local Similarity 53.8%; Pred. No. 1,7e-56;
 Matches 211; Conservative 62; Mismatches 109; Indels 10; Gaps 7;
 QY 1 MNRGVPRRH-LLVLTQLALLPATQGNKVVLGKGDVELTCTASQKKSIOFHKNSNOI 59
 DB 1 MCRATSLRRLLLLQLLSQLLAVTQGTLYLGKGESEALPCSSQKRTVFTWKFSDOR 60
 QY 60 KIINGOG-SFLTGS--PSKINDRADSRSLMDGNFPLIIKNKIENDSPYICVEDEKE 116
 DB 61 KIISQHGKGVILIRGSPSQF-DREPSKKGAMKSPFLINKLKMEDSQTICLENRKE 119
 QY 117 EVQLLVFGLTANSPDTHLLQGSLLTLTLDS--PPGSSPSPVQCRSPRKNIOGKTLVSQLE 175
 DB 120 EVELMWVKVTFSPETSLLOQSLLTLTLDSNSKVSNPFLTECHKKKGKVSQSLVSMNLR 179
 QY 176 LQDSGTCTCYLVNQKVEKIDIVLAFOKASSIVYKKEGEOVESFPPLAFVYEKLTGS 235
 DB 180 VQSDPFNQCVTYLLDQKKNMGMTLSVGFQSTAITAVKSGESAEFPPLFAEE--NGW 237
 QY 236 GELMWQNERASSKSWTFPLDKXKEVVKRVTQDPKLOMKKRLPLHLLTPOALPOVAGSG 295
 DB 238 GELMWKAKKDSFPQWTFSTIKNEVSVQSTKDLKQLKELTPLLTKIPVSLQFPGSG 297
 QY 296 NLTLALBAKTGKTLHQEVNLVVMRATQLOKNTLCEVWGPSPKMLSLKLENKAVSKRE 355
 DB 298 NLTLTLTD--KGTLHQEVNLVVMKVAQNLNLTLCGVWGPSPKMLTLKQENQEARVSEQ 355
 QY 356 KPVAVLNPEAGMWOCCLISDSGOVLLESNIYVL 387
 DB 356 KVVQVVAPEGTGLWQCLLSBGDKVMDSRIOVL 387
 RESULT 19
 GC3M_MOUSE

DR PIR; B02159; GIMSM.
 DR PDB; 15C8; 23-MAR-99.
 DR PDB; 1AE6; 18-MAR-98.
 DR PDB; 1CL7; 12-JAN-00.
 DR PDB; 1FL1; 06-FEB-01.
 DR PDB; 1F58; 29-DEC-99.
 DR PDB; 1K05; 24-JUL-02.
 DR PDB; 1KCR; 11-MAY-02.
 DR PDB; 25C8; 09-JUL-99.
 DR MGD; MGI:96446; 1gh-4.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003597; Ig_c1.
 DR InterPro; IPR003006; Ig_MHC.
 DR Pfam; PF00047; Ig; 2.
 DR SMART; SM00407; IGc1; 2.
 DR PROSITE; PSS0835; IG LIKE; 3.
 DR PROSITE; PSS0290; IG_MHC; 1.
 DR Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
 KM Alternative splicing; Transmembrane; 3D-structure.
 FT NON_TER 1 1
 FT DOMAIN 1 97 CH1.
 FT DOMAIN 98 110 HINGE.
 FT 111 217 CH2.
 FT DOMAIN 218 324 CH3.
 FT DISULFID 27 82
 FT DISULFID 102 102 INTERCHAIN (WITH A LIGHT CHAIN).
 FT DISULFID 104 104 INTERCHAIN (WITH A HEAVY CHAIN).
 FT DISULFID 107 107 INTERCHAIN (WITH A HEAVY CHAIN).
 FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
 FT DISULFID 138 198
 FT CAROHND 174 174 N-LINKED (GLCNAC. . .).
 FT TRANSMEM 340 357
 FT DOMAIN 358 393 POTENTIAL.
 SQ SEQUENCE 393 AA; 43386 MW; 4CC88343B7A1CE27 CRC64;

Query Match 27.5%; Score 940.5; DB 1; Length 393;
 Best Local Similarity 63.2%; Pred. No. 3,2e-53;
 Matches 165; Conservative 46; Mismatches 45; Indels 5; Gaps 2;

QY 388 PWSFVPEPKSCDKHTHTCP--CPAPELLGSGSVLFFPKPMDTLMIRTEPVTCVVVDV 445
 DB 87 PASSFVKVKKIVRCGCKPCICTVPEV---SSVFIFPKPKPDVLTITLTKKVCVVVDI 143
 QY 446 SHEDEVFKNVVDVEVHNNAKTPREQVNSTYVWVVLTVLHODMNLGKYEKKCVSNK 505
 DB 144 SKDDEVPQSMVVDVEVHTAQTQPREQFNSTFSSVSELPIMODMLNGKFKCRVNSA 203
 QY 506 ALPADIETKTSKAKQPREPOVYTLPSRDELTKNQVSLTCLVKGFYPSDIAVWESNGQ 565
 DB 204 AFPADIETKTSKTKGRPAPOVYTIPPKQGMKDKVSLTCLMTDFPFEDITVEMQNGQ 263
 QY 566 PENNYKTPPVLDSSGSEFVLSKLTVDKSRNQGVESGVMEALNHHYTKQSLSLSPG 625
 DB 264 PAENYKNTQPTMNTGSIYFVSKLVNQKSNMEAGNTFCVSLHGLNHHTEKSLSPG 323
 QY 626 LQLDETCAADQDGLGLMTT 646
 DB 324 LQLDETCAADQDGLGLMTT 344
 RESULT 21
 GC_RABIT
 ID_GC_RABIT STANDARD; PRT; 323 AA.
 AC P01870;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Ig gamma chain C region.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OC NCBI_TaxId=9986;

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84030930; PubMed=6313520;
 RA Bernstein K.E., Alexander C.B., Mage R.G.;
 RT "Nucleotide sequence of a rabbit IgG heavy chain from the recombinant
 RT F-1 haplotype.";
 RL Immunogenetics 18:387-397(1983).
 RN [2]
 RP SEQUENCE OF 1-128.
 RX MEDLINE=76135469; PubMed=1243651;
 RA Pratt D.M., Mole L.E.;
 RT "Sequence studies on the constant region of the Fd sections of rabbit
 RT immunoglobulin G of different allotype.";
 RL Biochem. J. 151:337-349(1975).
 RN [3]
 RP SEQUENCE OF 88-266 FROM N.A.
 RX MEDLINE=83299917; PubMed=6193512;
 RA Martens C.L., Moore K.W., Steinmetz M., Hood L., Knight K.L.;
 RT "Heavy chain genes of rabbit IgG: isolation of a cDNA encoding gamma
 RT heavy chain and identification of two genomic C gamma genes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 79:6018-6022(1982).
 RN [4]
 RP SEQUENCE OF 132-161.
 RX MEDLINE=70110015; PubMed=5461106;
 RA Fruchter R.G., Jackson S.A., Mole L.E., Porter R.R.;
 RT "Sequence studies of the Fd section of the heavy chain of rabbit
 RT immunoglobulin G.";
 RL Biochem. J. 116:249-259(1970).
 RN [5]
 RP SEQUENCE OF 129-131 AND 155-322.
 RA Hill R.L., Lebovitz H.E., Fellows R.E. Jr., Delaney R.;
 RL (In) Killeander J. (eds.);
 RL Gamma globulins, Nobel symp. 3, pp.109-127, Almqvist and Wiksell,
 RL Stockholm (1967).
 CC -1- MISCHELANEOUS: Ref.1 sequence has the D12 allotypic marker,
 CC 104-Thr, and the E14 marker, 185-Thr. Ref.3 has the D11 and E15
 CC markers and Ref.5 the E15 marker.
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; M16426; AAA31289.1; -.
 DR PIR; A91749; GHRB.
 DR HSSP; P01857; 1PCL.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003597; Ig_c1.
 DR InterPro; IPR003006; Ig_MHC.
 DR Pfam; PF00047; Ig; 3.
 DR SMART; SM00407; IGc1; 2.
 DR PROSITE; PSS0835; IG LIKE; 3.
 DR PROSITE; PSS0290; IG_MHC; 1.
 KM Immunoglobulin domain; Immunoglobulin C region; Repeat.
 FT NON_TER 1 1
 FT DOMAIN 1 96
 FT DOMAIN 114 213 IG-LIKE 1.
 FT DOMAIN 222 318 IG-LIKE 2.
 FT VARIANT 104 104 T -> M (IN D11 MARKER).
 FT VARIANT 185 185 T -> A (IN E15 MARKER).
 FT CONFLICT 48 48 N -> E (IN REF. 2).
 FT CONFLICT 71 71 V -> VPV (IN REF. 2).
 FT CONFLICT 144 144 Q -> E (IN REF. 3 AND 4).
 FT CONFLICT 173 173 N -> D (IN REF. 5).
 FT CONFLICT 187 187 Q -> E (IN REF. 3 AND 5).
 FT CONFLICT 201 201 N -> D (IN REF. 5).
 FT CONFLICT 218 218 Q -> E (IN REF. 5).
 FT CONFLICT 233 233 E -> Q (IN REF. 5).
 FT CONFLICT 246 246 N -> D (IN REF. 5).

```

FT FT CONFLICT 256 256 E -> G (IN REF. 5)
FT CONFLICT 250 250 N -> D (IN REF. 5)
FT CONFLICT 256 256 N -> D (IN REF. 5)
FT CONFLICT 280 280 Y -> W (IN REF. 5)
FT CONFLICT 284 284 N -> S (IN REF. 5)
SQ SEQUENCE 323 AA, 35404 MW, 69E8AA118D579A8B CRC64;

Query Match 27.0%; Score 922; DB 1; Length 323;
Best Local Similarity 69.3%; Pred. No. 3,8e-52;
Matches 169; Conservative 30; Mismatches 37; Indels 8; Gaps 2

OY 362 SNIKYLPMTSTVEEKSCKHTTCPCPAPELLGSGSVLPFPKPKDMLMISRTPEYTV 441
DB 87 TNYTKDKT---VADSTCK---PTCPPELLGSGSVLPFPKPKDMLMISRTPEYTV 138
OY 442 VDVSHEDPEVKNFYVDGVEVHNAAKTKREEOYNSTYVAVSVLTVLHODWLNKGEYCK 501
DB 139 VVDVSGQDDPEVQFTWYINNEQRTARAPRLBEQPFNSTIIVAGTLPITHQDWLRGKEFKCK 198
OY 502 VSNKALPAPIEKTISKAKGQPRPEOYVTLPPSRDELTKQVSLTCLVKGFYPSDIAVEME 561
DB 199 VHNKALPAPIEKTISKARQPLPEPKYITGPPREELISRSVSLTCLWIMINFPSPDISVEWE 258
OY 562 SNGQPENNYKTPPLVLDSDGSEFVYSKLTVDSDRMQGVFSCSVNHEALHNHYQKSL 621
DB 259 KNGKADENYKTPPAVLDSGSYFLYNKLSVPTSEWGRGVPTCSVMHEALHNHYQKSL 318
OY 622 LSPG 625
DB 319 RSPG 322

RESULT 22
GCAM_MOUSE
ID GCAM_MOUSE STANDARD; PRT; 399 AA.
AD P01865;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig gamma-2A chain C region, membrane-bound form.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82222190; PubMed=6283537;
RA Yamawaki-Kataoka Y., Nakai S., Miyata T., Honjo T.;
RT "Nucleotide sequences of gene segments encoding membrane domains of
RT Immunoglobulin gamma chain.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:2623-2627(1982).
CC -I- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Membrane-bound;
CC IsoId=P01865-1; Sequence=Displayed;
CC Name=Secreted;
CC IsoId=P01864-1; Sequence=External;
CC Note=Probably the major isoform;
CC -I- SIMILARITY: Contains 3 immunoglobulin-like domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J00471; AAB59661.1; ALT_INIT.
DR PIR; A02154; GZMSAM.
DR PDB; 1KB5; 08-APR-98.
DR PDB; 1YEB; 15-OCT-97.
DR MGD; MGI:96443; Igh-1.

```

DR		InterPro; IPR007110; Ig-like.
DR		InterPro; IPR003597; Ig_c1.
DR		InterPro; IPR003066; Ig_MHC.
DR	Pfam; PF00047; Ig_2.	
DR	SMART; SM00407; IGc1; 2.	
DR	PROSITE; PS00835; IG_LIKE; 3.	
DR	PROSITE; PS00290; IG_MHC; 1.	
KW	Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;	
KW	Transmembrane; Alternative splicing; 3D-structure; Repeat.	
FT	NON TER	1
FT	DOMAIN	6 98
FT	DOMAIN	121 220
FT	DOMAIN	229 325
FT	DISULFID	15 15
FT	DISULFID	27 82
FT	DISULFID	107 107
FT	DISULFID	110 110
FT	DISULFID	112 112
FT	DISULFID	144 204
FT	DISULFID	250 308
FT	TRANSMEM	346 363
FT	DOMAIN	364 399
FT	CARBOHYD	180 180
SEQ	SEQUENCE	399 AA; 44020 MM; 4C8138BBAED3FF0 CRC64;
Query Match	26.8%; Score 915.5; DB 1; Length 399;	
Best Local Similarity	49.6%; Pred. No. 1.3e-51;	
Matches 183; Conservative 48; Mismatches 79; Indels 59; Gaps 7		
QY	285 PQALPQVYAGSGNLTALFAKTKGKLHGEV-----NLVVMRAIQLOKNLTCCEWGPSPSKLM	339
DB	34 PEPVTLTNWSGSLSSGVHTFPFPAVLQSDLYTLLSSVTATVTSWPSQSITCNVAHPAS----	89
QY	340 LSLKLENKAAYKSREKEPVWLINPEAGMWQCCLSDSGOVULESNIKVLPTWSTVPKPKSC	399
DB	90 -----STVDKKIIP-----RG-----PT-----IKP----	106
QY	400 DKHTPCRP--CPABELLGSPVFLEPPPKXDTLMISRTPEVTVCVVDVSHDEPVKFNMY	457
DB	107 -----CPCKCRPALPLGGPFSEVFTEPPRIKQVLMISLPYTCVVVDVSDDPDVQISMF	161
QY	458 VDGVEYNNAKTRPREEOYNSTRVYSVLTVLHDMLNKEKYCKKSNKALLPAPIEKITSK	517
DB	162 VNNEVNHQAOTQTHREDYNSLTRVVSALPIQHODMMSKEFKCKVNNDLPAPIERITSK	221
QY	518 AKGOBREQVYTLPSPRDELTKNOVLSLCLKGFYPSDIIVAWESNGOPENNYKTTPYL	577
DB	222 PKGSVRABQVAVLPPPEBEEMTKQVTLCTMTDFEPEDIYVEMWTNGKTELNYKTEPYL	281
QY	578 DSDGSPFLYSKLTVDKSRMOOGNVFSCSVMEALAHNHYTOXLSLSPGLQDETCAEAQD	637
DB	282 DSDGSPFYMSKLRAREKKNWVERNSYSCSVVHEGLHNHTTSFSFRTPQLDDDVCAEAQD	341
QY	638 GELDGLMTT 646	
DB	342 GELDGLMTT 350	
RESULT 23		
GCD_CAVPO	STANDARD;	PRT; 329 AA.
ID	GCD_CAVPO	
AC	P01862;	
DT	21-JUL-1986 (Rel. 01, Created)	
DT	21-JUL-1986 (Rel. 01, Last sequence update)	
DT	15-JUL-1999 (Rel. 38, last annotation update)	
DS	Ig gamma-2 chain C region.	
OS	Cavia porcellus (Guinea pig).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OX	Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.	
NCBI_TaxId=10141;		
RN	[1]	
SEQUENCE OF 1-3.		
RA	Trieschmann T.M.;	

```

RL Submitted (APR-1975) to the PIR data bank.
RN
RP SEQUENCE OF 4-68.
RX MEDLINE=71058471; PubMed=5538606;
RA Birhetein B.K., Husain Q.Z., Cebra J.J.;
RT "Structure of heavy chain from strain 13 guinea pig
RT immunoglobulin-G(12). 3. Amino acid sequence of the
RT half-cysteine joining heavy and light chains.";
RL Biochemistry 10:18-25(1971).
RN
RP SEQUENCE OF 69-133 AND 312-329.
RX MEDLINE=71058486; PubMed=5538616;
RA Turner K.J., Cebra J.J.;
RT "Structure of heavy chain from strain 13 guinea pig
RT immunoglobulin-G(12). 11. Amino acid sequence of the
RT and hinge region cyanogen bromide fragments.";
RL Biochemistry 10:9-17(1971).
RN
RP SEQUENCE OF 134-226.
RX MEDLINE=75036072; PubMed=4429665;
RA Tracey D.E., Cebra J.J.;
RT "Primary structure of the CH2 homology region from guinea pig IgG2
RT antibodies.";
RL Biochemistry 13:4796-4803(1974).
RN
RP SEQUENCE OF 227-311.
RX MEDLINE=75036073; PubMed=4609467;
RA Trischmann T.M., Cebra J.J.;
RT "Primary structure of the CH3 homology region from guinea pig IgG2
RT antibodies.";
RL Biochemistry 13:4804-4811(1974).
RN
RP DISULFIDE BONDS.
RX MEDLINE=71058474; PubMed=4922544;
RA Oliveira B., Lamm M.E.;
RT "Interchain disulfide bridges of guinea pig gamma-2-immunoglobulin.";
RL Biochemistry 10:26-31(1971).
CC -1- MISCELLANEOUS: This chain was isolated from pooled serum of strain
CC 13 inbred guinea pigs.
DR PIR; A94553; G2GP.
DR HSSP; P01842; 7RAB.
DR InterPro; IPR007110; Ig-1-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 2.
DR SMART; SM00407; IGc1; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; glycoprotein.
FT NON_TER 1 1
FT DISULFID 16 16 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 28 79
FT DISULFID 105 105 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 107 107 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 110 110 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 142 202 INTERCHAIN (WITH A HEAVY CHAIN).
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .).
FT DISULFID 248 308
SQ SEQUENCE 329 AA; 36074 MW; 5D231B7164D1FBA9 CXC64;

Query Match 26.6%; Score 906.5; DB 1; Length 329;
Best Local Similarity 58.1%; Pred. No. 3.8e-51;
Matches 176; Conservative 31; Mismatches 57; Indels 39; Gaps 4;
QY 325 NLTCGVMPRTSPKMLSLKLENKAKVSKREKPVVNLPEAGMOCCLISDGYLLSNI 384
DB 64 SLTGWVTVBSSOKATCNVAHAPASSTKVKTVETPI----- 97
QY 385 KVLPTWSTVPPPKSCDKTHTCPAPRLGSPVFLFPKPKDTMTSRTPEVTCVVD 444
DB 98 -----RTP-ZRPBC-----TCKPCPPNLDGSPSVFIRPPKDTLMSLTLPVTCVVD 146
QY 445 VSHEDPEVKFMWYDGVENHAKTKPREEOYNSTYVSVLTVLHODWLNGEKYCKVSN 504

```

```

DB 147 VSODEPEVQFTWPEVDNKPVENAETKPEVEQYNTTFVRESVLPLOHODMLGKEFKCYVN 206
QY 505 KALPAPIEKTISAKAGQPREPOVYTLPPSRDELTJKNVSLTCLVKGYPADIVEMSSNG 564
DB 207 KALPAPIEKTISKTGAPRPMDVYTLPPSRDELSKSVSYTCLINFPADIVEMASNR 266
QY 565 QP--ENNYKTPPVLDSDGSFPLYSKLTVDKSRMOQGNVSCSYMEBALNHYTKSLSL 622
DB 267 VPVSEKEYKNTPTPEADDSYFLYSKLTVDKSAAMDQCTVTCGYMEBALNHYTKSLISR 326
QY 623 SPG 625
DB 327 SPG 329

RESULT 24
GCBM_MOUSE STANDARD; PRT; 405 AA.
ID GCBM_MOUSE
AC P01867;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig gamma-2B chain C region, membrane-bound form.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN
RP SEQUENCE OF 335-405 FROM N.A.
RX MEDLINE=8222190; PubMed=628537;
RA Yamawaki-Kataoka Y., Nakai S., Miyata T.;
RT "Nucleotide sequences of gene segments encoding membrane domains of
RT immunoglobulin gamma chains.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:2623-2627(1982).
RN
RP SEQUENCE OF 335-378 FROM N.A.
RX MEDLINE=82115295; PubMed=6799207;
RA Rogers J., Choi E., Souza L., Carter C., Word C.J., Kuehl M.,
RA Eisenberg D., Wall R.;
RT "Gene segments encoding transmembrane carboxyl termini of
RT immunoglobulin gamma chains.";
RL Cell 26:19-27(1981).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Membrane-bound;
CC IsoId=P01867-1; Sequence=Displayed;
CC Name=Secreted;
CC IsoId=P01866-1; Sequence=External;
CC Note=May be the major isoform;
CC -1- PPM: DISULFIDE BONDS BY HOMOLOG WITH OTHER IG GAMMA CHAINS.
CC -1- MISCELLANEOUS: The sequence of residues 1-335 is assumed to be
CC identical with the corresponding region of the secreted form.
CC -1- MISCELLANEOUS: The a allele sequence is shown.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announce/
CC or send an email to license@ebi.ac.uk).

```

DR PROSITE; PS50835; IG_LIKE; 3.
 DR PROSITE; PS00290; IG_MHC; 1.
 KW Immunoglobulin domain; Immunoglobulin C region; Transmembrane;
 KW Alternative splicing; 3d-structure; Repeat.
 FT NON_TER 1 1
 FT DOMAIN 6 98 IG-LIKE 1.
 FT DOMAIN 127 226 IG-LIKE 2.
 FT DOMAIN 235 331 IG-LIKE 3.
 FT DISULFID 15 15 INTERCHAIN (WITH A LIGHT CHAIN).
 FT DISULFID 27 82 INTERCHAIN (WITH A LIGHT CHAIN).
 FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
 FT DISULFID 112 112 INTERCHAIN (WITH A HEAVY CHAIN).
 FT DISULFID 115 115 INTERCHAIN (WITH A HEAVY CHAIN).
 FT DISULFID 118 118 INTERCHAIN (WITH A HEAVY CHAIN).
 FT DISULFID 150 210 INTERCHAIN (WITH A HEAVY CHAIN).
 FT DISULFID 256 314
 FT TRANSMEM 352 369
 FT DOMAIN 370
 SQ SEQUENCE 405 AA; 44330 MW; 8983CF0A9B6D49FA CRC64;
 POTENTIAL.
 CYTOPLASMIC (POTENTIAL).

Query Match Best Local Similarity 25.9%; Score 884.5; DB 1; Length 405;
 Matches 174; Conservative 53; Mismatches 91; Indels 49; Gaps 4;

QY 285 PQLPQYAGSGNLTLLAE-----KTGKLHGVNLVVRATOLQKLTCEVWGPTSPKLM 339
 DB 34 PSVSYVTNMGSSLSVHTFPALLOGLYTMSSSVTPSPQVTCVSAHPAS-STT 92
 QY 340 LSLKLENKAAVSKREKPVVNLPEAGWQCLLSDSGQVLLSNIKVLPTWSTPVEPKSC 399
 DB 93 VKKLE-----PSCPSTINP-----C 109
 QY 400 DKTHNCPGAPBELLGSPVFLPPPKKDTLMISRTPEVTCVVDVSHEDPEVKRWYD 459
 DB 110 PCKCKHCKPAPLLEGSPVFLFPKPKVLMISLTPKVTCAVVDSEDDPVQISWPN 169
 QY 460 GVEVNAKTPREEQYSTYRVVSLVTLHODMLNKEYCKVSKNKPAPIEKTISKAK 519
 DB 170 NVEVHTAQVTHREDNSTIRVSTLPIQHODMSKEKCKVNNKDLPSPIERTISKK 229
 QY 520 GQPREQVYTLPPSRDELTKNOVSLTCLVKGFPYPSDIAVWESNQGPENNYKTPPVLD 579
 DB 230 GLVRAQVYILPPPAEQLSRKDVSLTCLVGVFPGDISVEMTNGHTEENYKDTAPVLD 289
 QY 580 DGSFPLSLTLYDKSMQGNVPSGCVMEALHNHTQKSLSPQLDDEFCARQGE 639
 DB 290 DGSYFLYSLLNKKTSKMEKTDSPSCNVRHEGLKNYYLKXITISRPGLDLDDICAEKQGE 349
 QY 640 LDGLWTT 646
 DB 350 LDGLWTT 356

RESULT 25

GC3_MOUSE STANDARD; PRT; 329 AA.
 AC P22536;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 16 gamma-3 chain C region, secreted form.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OK NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85027161; PubMed=6092053;
 RA Wele J.A., Word C.J., Rimm D., Der-Balan G.P., Martinez H.M.,
 RA Tucker P.W., Blatner F.R.;
 RT "Structural analysis of the murine IgG3 constant region gene.";
 RT EMBO J. 3:2041-2046(1984).
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; J00451; -; NOT_ANNOTATED_CDS.

DR PIR; B02156; G3MSC.

DR HSSP; P01857; 1FC1.

DR InterPro; IPR007110; IG-1like.

DR InterPro; IPR003597; IG_C1.

DR InterPro; IPR003006; IG_MHC.

DR Pfam; PF00047; Ig3.3.

DR SMART; SM00407; IGC1.2.

DR PROSITE; PS50835; IG_LIKE; 3.

DR PROSITE; PS00290; IG_MHC; 1.

KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
 KW Transmembrane; Alternative splicing.

FT NON_TER 1 1
 FT DOMAIN 1 97 CH1.
 FT DOMAIN 98 113 HINGE.
 FT DOMAIN 114 223 CH2.
 FT DOMAIN 224 327 CH3.

SQ SEQUENCE 329 AA; 36228 MW; F45827174182BAD6 CRC64;

Query Match Best Local Similarity 24.9%; Score 849.5; DB 1; Length 329;
 Matches 154; Conservative 42; Mismatches 48; Indels 21; Gaps 2;

QY 382 SNIKVLPMTWSTPVEPKSCDKTH-----TCPP--CPABELGSPVFL 420
 DB 64 SLLVTPSSSTMSQVQCNVAHPASKTELIRIEPRIPKSTPSSGCPGNILGSPVFL 123
 QY 421 LFPKPKDTLMISRTPEVTCVVDVSHEDPEVKRWYDGVVNAKTPREEQYSTYR 480
 DB 124 IFPPKPKALMISLTPKVTCAVVDSEDDPVHWSFNDKNEVHTAMQPREAQVNSTFR 183
 QY 481 VSVSLTTLHODMLNKEYCKVSKNKPAPIEKTISKAKGQPREQVYTLPPSRDELTKN 540
 DB 184 VVSALPIQHODMNGEKEKCKVNNKDLPSPIERTISKKGAQVYTLIPPREQMSKK 243
 QY 541 QVSLTCLVKGFPYPSDIAVWESNQGPENNYKTPPVLDSDSFLYSLLTLYDKSMQGN 600
 DB 244 KVSLLTCLVTFPSEASIVEMERNGELQDYKNTPEILDSDYFLYSLLTLYDTSWLOGE 303
 QY 601 VPSGCVMEALHNHTQKSLSPG 625
 DB 304 IFTCSVMEALHNHTQKSLSPG 328

RESULT 26

GCB_RAT STANDARD; PRT; 333 AA.
 AC P20761;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE 16 gamma-2B chain C region.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OK NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89232738; PubMed=3149946;
 RA Brueggemann M.;
 RT "Evolution of the rat immunoglobulin gamma heavy-chain gene family.";
 RT Gene 74:473-482(1988).
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
 DR PIR; PS0018; PS0018.
 DR HSSP; P01842; 7FAB.

```

DR InterPro: IPR007110; IG-like.
DR InterPro: IPR003597; IG_C1.
DR InterPro: IPR003006; IG_MHC.
DR Pfam: PF00407; Ig: 3.
DR SMART: SM00407; IGc1. 2.
DR PROSITE: PS00835; IG-LIKE; 3.
DR PROSITE: PS00290; IG_MHC; 1.
DR Immunoglobulin domain; Immunoglobulin C region; Repeat.
FT NON TER 1 1
FT DOMAIN 6 96 IG-LIKE 1.
FT DOMAIN 124 223 IG-LIKE 2.
FT DOMAIN 232 328 IG-LIKE 3.
FT DISULFID 15 15 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 27 80 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 112 112 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 115 115 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 147 207 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 253 311
SQ SEQUENCE 333 AA; 36497 MW; 55F8B64D48D460A6 CRC64;

Query Match 24.7%; Score 844; DB 1; Length 333;
Best Local Similarity 61.7%; Pred. No. 3.9e-47;
Matches 153; Conservative 34; Mismatches 51; Indels 10; Gaps 1;

388 PWSSTPVPK-----SCDKHTCTPCPCAPELLGSPVFLPFPKXDTLMISRTPE 437
DB PASSTKVKYKVRNRNGIGHKCPTCTCHKCPVPELLGSPVFLPFPKXDTLMISQNAK 144
QY 438 VTCVVVDVSHEDPEVKFMYVDGVEVNAKTRPEEQYNSITVRVSVLTVLHQMUNGKE 497
DB 145 VTCVVVDVSEEDPQVQFMPVNVVEVHTAQTPREEQYNSITFRVVSALPIQHDMWSGKE 204
QY 498 YKCKSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIA 557
DB 205 FKCKRNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIA 264
QY 558 VEMESNGQPENNYKTPPVLDSDGSFPLYSKLTVDKSRMQQGNVSCSMHEALHNNHYTQ 617
DB 265 VEMISNGIIEKQYKNTPEVMDSDGSFPLYSKLTVDKSRMQQGNVSCSMHEALHNNHYTQ 324
QY 618 KSLSLSPG 625
DB 325 KSLSLSPG 332

RESULT 27
GCL_RAT STANDARD; PRT; 326 AA.
ID GCL_RAT AC P20759;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig gamma-1 chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89232738; PubMed=3149946;
RA Brueggemann M.;
RL "Evolution of the rat immunoglobulin gamma heavy-chain gene family.";
RL Gene 74:473-482(1988).
DR HSSP; P01842; 7FAB.
DR InterPro: IPR007110; IG-like.
DR InterPro: IPR003597; IG_C1.
DR InterPro: IPR003006; IG_MHC.
DR Pfam; PF00407; Ig: 3.
DR SMART; SM00407; IGc1; 2.
DR PROSITE; PS00835; IG-LIKE; 3.

```

```

DR PROSITE: PS00290; IG_MHC; 1.
DR Immunoglobulin domain; Immunoglobulin C region; Glycoprotein.
FT NON TER 1 1
FT DOMAIN 98 97 CH1.
FT DOMAIN 113 219 HINGE.
FT DOMAIN 220 326 CH2.
FT DISULFID 27 82 CH3.
FT DISULFID 102 102 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 111 111 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 140 200
FT CARBOHYD 176 176
SQ SEQUENCE 326 AA; 35946 MW; 013BAB45EFA9B9DA CRC64;

Query Match 24.3%; Score 828; DB 1; Length 326;
Best Local Similarity 55.1%; Pred. No. 4.1e-46;
Matches 150; Conservative 48; Mismatches 48; Indels 26; Gaps 4;

375 SGCVLLESNIKVLPTWSTPVPKSCDKTH-----TCCPCAPPELLGG 416
DB 59 SGLYTLTSSVTV-PSSTWPSQVTCVNAHPASSTKDKKIVPRNCGDCKPC---ICTG 113
QY 417 ---PSVFLPFPKXDTLMISRTPEVTCVVVDVSHEDPEVKFMYVDGVEVNAKTRPEE 473
DB 114 SEVSVVIFPFPKXDTLTLTKVTCVVVDISQDDPEVAFSMVDVEVHTAQTRPEE 173
QY 474 QYNSTRVSVLTVLHQMUNGKEVYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPS 533
DB 174 QFNSTFRSVSELDPLHDMUNGRTFRCKVTSAAFPSPIEKTISKPEGRTPVPHYTWSP 233
QY 534 RDELTKNQVSLTCLVKGFYPSDIAVEMESNGQPENNYKTPPVLDSDGSFPLYSKLTVDK 593
DB 234 KEEMTQNEVSIITCVKGFYPPDIYVEMQNGQPENNYKNTPTMDTGSIFLYSKLVNKK 293
QY 594 SRMQGNVSCSMHEALHNNHYTQKSLSLSPG 625
DB 294 EKMQGNTFTCSVLHDEGLHNHNTEKSLSLSPG 325

RESULT 28
GCL_MOUSE STANDARD; PRT; 324 AA.
ID GCL_MOUSE AC P01868;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig gamma-1 chain C region secreted form.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=80045036; PubMed=115593;
RA Honjo T., Obara M., Yamawaki-Kataoka Y., Kataoka T., Kawakami T.,
RA Takahashi N., Mano Y.;
RT "Cloning and complete nucleotide sequence of mouse immunoglobulin
RT gamma 1 chain gene.";
RL Cell 18:559-568(1979).
RN [2]
RP SEQUENCE OF 76-324 FROM N.A. (MYELOMA PROTEIN MOPC 31C).
RX MEDLINE=80202553; PubMed=6769752;
RA Obara M., Yamawaki-Kataoka Y., Takahashi N., Kataoka T., Shimizu A.,
RA Mano Y., Seidman J.G., Peterlin B.M., Leder P., Honjo T.;
RT "Immunoglobulin gamma 1 heavy chain gene: structural gene sequences
RT cloned in a bacterial plasmid.";
RL Gene 9:87-97(1980).
RN [3]
RP SEQUENCE OF 70-322 FROM N.A. (MYELOMA PROTEIN MOPC 21).
RX MEDLINE=80012837; PubMed=113776;

```


Query Match	Similarity	Score	DB 1	Length	324
Best Local	60.0%	Pred. No. 5e-46			
Matches 144	Conservative	46	Mismatches	45	Indels 5; Gaps 2

Qy	388	PTWSTPVPSPSCDTHXCP--CAPAPLLGGPSVFLPPPKKOTLMSRTPPEVTCVVADV	445
Db	87	PASTTKVDKTIKVPDCCGCKPCTCIVEP---SSVIFPPPKKOVLTITLTPKTCVVVDI	143
Qy	446	SHEDPEVKFNMYVDGVEVHNAKTKPREEOYNSTYRVVSVLTVLHQMVLNGEKYCKVSNK	505
Db	144	SKDPDPEVQFSVFDVDVHTKQTOPRDEQFSTFRSVSELPIRMQMDLNGEFKCRVNSA	203
Qy	506	ALPAPIKKTISKAGQREPOVYVTLPSRSBELTNQVSLTCLVYKGFPSDLAVWESNGQ	565
Db	204	AFPAPIKKTISKTRGPKAPVYVTLPPPKOMADKSLTCMTDPPEDITVEMQNGQ	263
Qy	566	PENNYKTPPTPLDSDGSFFLVSKLTVKSRNQOANVSCSVMEHALNHYTKOXLSPG	625
Db	264	PAENYKTOPPLMNTNGSYFYISKLVNOKSKNWEAGNTTCSVLEHGLNHHTEKSLSHSPG	323
RESULT 29			
GCC_RAT			
AC	ID	GCC_RAT	STANDARD; PRT; 329 AA.
AC	P20762;		
DT	01-FEB-1991	(Rel. 17, Created)	
DT	01-FEB-1991	(Rel. 17, Last sequence update)	
DT	15-JUL-1999	(Rel. 38, Last annotation update)	
DE	Ig gamma-2C chain C region.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
NC	NCBI_TaxID=10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=88166903; PubMed=3127222;		
RA	Bruggemann M., Delmasro-Galfre P., Waldmann H., Galabi F.;		
RT	"Sequence of a rat immunoglobulin gamma 2c heavy chain constant		
RL	region cDNA: extensive homology to mouse gamma 3.";		
RL	Eur. J. Immunol. 18:317-319 (1988).		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; X07189; CA30169.1; .		
DR	PIR; S00847; S00847.		
DR	HSSP; P01842; 7FAB.		
DR	InterPro; IPR007110; Ig-like.		
DR	InterPro; IPR003597; Ig cl.		
DR	InterPro; IPR003006; Ig_MHC.		
DR	Pfam; PF00047; Ig; 2.		
DR	SMART; SM00407; Igcl; 2.		
DR	PROSITE; PS50835; IG_LIKE; 3.		
DR	PROSITE; PS00290; IG_MHC; 1.		
KW	Immunoglobulin domain; Immunoglobulin C region.		
FT	NON TER	1	
FT	DOMAIN	1	97
FT	DOMAIN	98	113
FT	DOMAIN	114	222
FT	DOMAIN	223	329
FT	DISULFID	15	15
FT	DISULFID	27	82
FT	DISULFID	111	111
FT	DISULFID	113	113
FT	DISULFID	143	203
FT	DISULFID	249	307
SEQ	SEQUENCE	329 AA; 36571 MW; 5FCDB7B933850773 CRC64;	
Query Match			
Base Local Similarity			
Matches 152; Conservative 46; Mismatches 51; Indels 23; Gaps 4;			
23.9%; Score 817.5; DB 1; Length 329;			
55.9%; Pred. No. 1.9e-45;			

```

OY 375 SGQVLLSNINX-LEPTWSTPVEPKSCDKTHTC-----PP---CPABEL 413
DB 59 SGLYLTSSSVTPSSTWSS--QTVCSVAHPATKSNLIKRIEPRRKRPRPTDICSCDN 116
OY 414 LGSPSVFLPPPKKDTLMISRTPEYTCVVVDVSHEDPEVKFMVYDGVENAKTKPR 473
DB 117 LGRPSVFLPPPKKDTLMITLTPKTCVVVDVSEEPVQPSWFDVNRVFTAQTPHEE 176
OY 474 QYNSTYRVSVLTVLHQMPLNGKCKYKSNKALPAPIEKTISKAKGQPREPQVYTLPPS 533
DB 177 QLNQFRVYVSTLTHIQDQMSGKCKYKSNKALPAPIEKTISKAKGQPREPQVYTLPPS 236
OY 534 RDELTKNOVSLTCLVKGFPYPSDIAVEMESNGQPENNYKTPPVLDSDGSFFLYSKLTVDK 593
DB 237 REQSKNKVSLTCWTSFYSPASISVEMERNGELQEDYKNTLPVLDSDGSFFLYSKLTVDK 296
OY 594 SRMOGNVFSQSVMEALHNHYTQKSLSLSPG 625
DB 297 DSMWGDYTCVSVHEALHNHTQKNLSRSPG 328

RESULT 30
GCAB_MOUSE STANDARD; PRT; 335 AA.
AC P01864;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig gamma-2a chain C region secreted form (B allele).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=82037861; PubMed=6170065;
RA Schreier P.H., Bochevelli A.L.M., Mueller-Hill B., Baltimore D.;
RT "Multiple differences between the nucleic acid sequences of the
IGG2a and IGG2ab alleles of the mouse.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:4495-4499(1981).
RN [2]
RP SEQUENCE.
RX MEDLINE=82037777; PubMed=6794027;
RA Dognin M.J., Lauwereys M., Strosberg A.D.;
RT "Multiple amino acid substitutions between murine gamma 2a heavy
chain Fc regions of Ig1a and Ig1b allotypic forms.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:4031-4035(1981).
CC -1- SUBCELLULAR LOCATION: Secreted (Potential).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Secreted;
CC IsoId=P01864-1; Sequence=Displayed;
CC Note=Probably the major isoform;
CC Name=Membrane-bound;
CC IsoId=P01865-1; Sequence=External;
CC -1- MISCELLANEOUS: The sequence differs from that of the a allele,
from BALB/c mice, at 15 of the positions.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by, and for commercial
entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; J00479; -; NOT_ANNOTATED_CDS.
CC PIR; A02153; G2MSAB.
CC PDB; 1BOG; 23-MAR-99.
CC PDB; 1HH6; 26-JAN-01.

```

```

DR PDB; 1HH9; 24-JUL-03.
DR PDB; 1HI6; 08-FEB-01.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig cl.
DR InterPro; IPR003066; Ig_MHC.
DR Pfam; PF00047; Ig; 2.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PS50835; IG-LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; Alternative splicing;
KW 3D-structure; Repeat.
FT NON TER 1
FT DOMAIN 6 98 IG-LIKE 1.
FT DOMAIN 126 225 IG-LIKE 2.
FT DOMAIN 234 330 IG-LIKE 3.
SO SEQUENCE 335 AA; 36596 MW; FA3382792CBB136C CRC64;

Query Match 23.9%; Score 816; DB 1; Length 335;
Best Local Similarity 45.7%; Pred. No. 2,5e-45;
Matches 158; Conservative 52; Mismatches 86; Indels 50; Gaps 3;

OY 285 POALPOVYAGSGNLTIALEA-----KTGKLEHENVLYVMRATQLOKNTCEVWGPTSPKLM 339
DB 34 PEVYTLTWNSSGLSSGVHTPTALLQSLYTLSSVYTSNTWPSQITTCNVHPAS---- 89
OY 340 LSLKENKEAKVSKREKPVVWLNPEAGMOCLLSDQVLLSNIKVLPWSTPVEPKSC 399
DB 90 -----SRKVDKIEP-----RVPTQNPC 108
OY 400 DKHTCPCCAPBELGSPVFLPPPKKDTLMISRTPEYTCVVVDVSHEDPEVKFMVYD 459
DB 109 PPHGRVPVCAAPDLGSPVFLPPPKTKDVLMTSLSMVTCVVVDSEDDPDVQISWPN 168
OY 460 GVEVHNKTKPREQVSTYRVSVLTVLHQMPLNGKCKYKSNKALPAPIEKTISKAK 519
DB 169 NVEVHTAQIQTHREDYNTLRVVSALPIQDQMSGKCKYKSNKALPAPIEKTISKAK 228
OY 520 GQPREPQVYTLPPSRDELTKNOVSLTCLVKGFPYPSDIAVEMESNGQPENNYKTPPVLD 579
DB 229 GPVYAPQVYVLPPEAPEMTKERSLTCTMIGFLPAEIAVDMWTSNGRTQVYKNTATVLD 286
OY 580 DGSFFLYSKLTVDKSRWQGNVFSQSVMEALHNHYTQKSLSLSPG 625
DB 289 DGSYFMYSKLRVOKSTWERSLPAQSVVHEALHNHTTQKTSIRSLG 334

RESULT 31
GCCA_MOUSE STANDARD; PRT; 330 AA.
AC P01863;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig gamma-2a chain C region, A allele.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81076554; PubMed=6777755;
RA Slikorav J.-L., Aufray C., Rougeon F.;
RT "Structure of the constant and 3' untranslated regions of the murine
Balb/c gamma 2a heavy chain messenger RNA.";
RL Nucleic Acids Res. 8:3143-3155(1980).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=81198976; PubMed=6262729;
RA Yamawaki-Kataoka Y., Miyata T., Honjo T.;
RT "The complete nucleotide sequence of mouse immunoglobulin gamma 2a gene
and evolution of heavy chain genes: further evidence for intervening
Nucleic Acids Res. 9:1365-1381(1981).

```

```

RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=81223894; PubMed=6787604;
RA Olio R., Aufray C., Mochamps C., Rougeon F.;
RT "Comparison of mouse immunoglobulin gamma 2a and gamma 2b chain genes
RT suggests that exons can be exchanged between genes in a multigenic
RT family";
RL Proc. Natl. Acad. Sci. U.S.A. 78:2442-2446(1981).
RN [4]
RP MYELOMA PROTEIN MOPC 173.
RX MEDLINE=74175517; PubMed=4831970;
RA Bourgois A., Fougereau M., Rocca-Serra J.;
RT "Determination of the primary structure of a mouse IgG2a
RT immunoglobulin:amino-acid sequence of the Fc fragment. Implications
RT for the evolution of immunoglobulin structure and function.";
RL Eur. J. Biochem. 43:423-435(1974).
RN [5]
RP DISULFIDE BONDS.
RX MEDLINE=73056887; PubMed=4565406;
RA de Preval C., Fougereau M.;
RT "Determination of the primary structure of a mouse gamma G2a
RT immunoglobulin. Identification of the disulfide bridges.";
RL Eur. J. Biochem. 30:452-462(1972).
CC -I- SIMILARITY: Containe 3 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ibt-sib.ch/announce/
CC or send an email to license@ibt-sib.ch).
CC -----
DR EMBL; V00798; CAA24178.1; -.
DR PIR; A02152; G2MSA.
DR PDB; 1E4W; 12-JUL-01.
DR PDB; 1E4X; 12-JUL-01.
DR PDB; 1KNU; 06-MAY-99.
DR InterPro; IPR007110; IG-1Lke.
DR InterPro; IPR003597; IG_C1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig_2.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; 3D-structure; Repeat.
FT FT NON TER 1 1
FT DOMAIN 6 98
FT FT 121 220
FT DOMAIN 229 325
FT FT 15 15
FT DISULFID 27 82
FT DISULFID 107 107
FT DISULFID 110 110
FT DISULFID 112 112
FT DISULFID 144 204
FT DISULFID 250 308
FT MOD RES 330 330
SQ SEQUENCE 330 AA; 36389 MW; B64361C545A6864 CMC64;
Query Match 23.9%; Score 814.5; DB 1; Length 330;
Best Local Similarity 47.4%; Pred. No. 3e-45;
Matches 165; Conservative 47; Mismatches 77; Indels 59; Gaps 7;
Oy 285 POLPQVAGSNTLTALAEAKTGKLGHEV-----NLVVMKATQLOKNTLCEVWGPTSPKLM 339
Db 34 PEPVLTITNSGSLSGVHTPPAVLQSDLYTLSSSVTVYTSWPSQISITCVNAPAS---- 89
Oy 340 LSLKLENKAVYSKREKPEVWVILNPEAGMOCCLSDSGVLLLESNIKVLPTSTVEPEKSC 399
Db 90 -----STVVDKIEP-----RG-----PT-----IKR--- 106
Oy 400 DKHTTCPP--CPAPELLGGSVFLFPKKPKQDTLMISRTPEVTCVVVDVSHEDPEVKENWY 457

```

```

Db      107  -----CPCCKCAPVLLGSPSVFLFPPEKIKVLMISLSPITVCVVVSDDDPVOQISWF 161
QY      458  VDGVEVHNAAKKPREEOQNSTYTRVSVVLTVLHODMLNGKEYKCKVSNKALPAPIEKTSK 517
Db      162  VNNVEVHTAQOTREDYNSLTLRVVSLPFIQHODMSKREKCKVNNKDLPAPIERTISK 221
QY      518  AKGQPREPOVYTLTPPSRDELTKNOVSLTCLVKGFPSPDIAVEMWNSQGPENNYKTPPV 577
Db      222  PKGSVRAPQVYVLLPPPEEEMTKKQVTLTCMTVDPEMEDIVYEMTNNGKTELNTYKTEPVL 281
QY      578  DSDGSFYLXSLTYTDKSRWQGNVPSGSGVMEALHNHTQSLSLSPG 625
Db      282  DSDGSYFYMSKLRVEKKNNWVRNYSVSGSVHEGLHNHTTKTSFSRTPG 329

RESULT 32
GCA_RAT
ID_GCA_RAT      STANDARD;      PRT;      322 AA.
AC      P20760;
DT      01-FEB-1991 (Rel. 17, Created)
DT      01-FEB-1991 (Rel. 17, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Ig gamma-2A chain C region.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxId=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=89232738; PubMed=3149946;
RA      Bruggemann M.;
RT      "Evolution of the rat immunoglobulin gamma heavy-chain gene family.";
RL      Gene 74:473-482(1988).
CC      -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M13804; AAA41376.1; ALT_INIT.
DR      PIR; PS0019; PS0019.
DR      HSSP; P01842; 7FAB.
DR      InterPro; IPR007110; Ig-like.
DR      InterPro; IPR003597; Ig_c1.
DR      InterPro; IPR003006; Ig_MHC.
DR      Pfam; PF00047; Ig; 2.
DR      SMART; SM00407; Igcl; 2.
DR      PROSITE; PS50835; IG_LIKE; 3.
DR      PROSITE; PS00290; IG_MHC; 1.
KW      Immunoglobulin domain; Immunoglobulin C region; Repeat.
FT      NON TER      1
FT      DOMAIN      6 98      IG-LIKE 1.
FT      DOMAIN      115 212      IG-LIKE 2.
FT      DOMAIN      221 317      IG-LIKE 3.
FT      DISULFID      27 82
FT      DISULFID      102 102      INTERCHAIN (WITH A HEAVY CHAIN).
FT      DISULFID      105 105      INTERCHAIN (WITH A HEAVY CHAIN).
FT      DISULFID      107 107      INTERCHAIN (WITH A HEAVY CHAIN).
FT      DISULFID      136 196      INTERCHAIN (WITH A HEAVY CHAIN).
FT      CARBOHYD      242 300
FT      FT      172 172      N-LINKED (GLCNAC... ) (POTENTIAL).
SQ      SEQUENCE      322 AA; 35186 MW; E8EA136A9DE01EDB CRC64;

Query March      23.3%; Score 795; DB 1; Length 322;
Best Local Similarity 56.4%; Pred. NO.5.2e-44;
Matches 150; Conservative 39; Mismatches 59; Indels 18; Gaps 5;
375 SGOVLEBSNIKV-LPTWST-----PVEPKSGCDK---THNCP--CPAPELLGSPV 419

```

```

Db      59 SGLYLTSSVTPSSTWSSQAVTCVAVAPASSTKYDKKIVPRECNPCQCTSEV---SSV 115
Oy      420 FLFPKPKMDTLMISTREPVTCVVVDVSHEDPEVFNNVVDVGVHNNKTRKREQVNSTY 479
Db      116 FLFPKPKMDVLTTLTPKVTCTVVDVDSQDEVRFSWTFIDVEVHTAQTAPKQKSTL 175
Oy      480 RVASVLTVLHODWLNKGEKCKVSKNKAUPAPIEKTISKAKQCPREPQVYTLPPSRDELTK 539
Db      176 RVSSELPIVHMDWLNKGTFCCKVSKNKAUPAPIEKTSISKSEPTPGQVYTTAPKREMTQ 235
Oy      540 NQVSLTCLVKGFPYSDIAVEMESNQCPENNYKTTTPVLDSQGSFPLSKLTVDKSRWQOG 599
Db      236 SQVSTICVWKGFYPPDITVEKWMKMGQPOENYKNTPTMTDTSYFLYSKLVNKKETWQOG 295
Oy      600 NVFSGSVMEALHNNHYTKSLSPG 625
Db      296 NTFCTSVLHBLHNNHTEKSLSHSPG 321

RESULT 33
GCB_MOUSE STANDARD; PRT; 336 AA.
AC P01866;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig gamma-2b chain C region secreted form.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A. (ALLELE A).
RX MEDLINE=80120716; Pubmed=6766534;
RA Yamawaki-Kataoka Y., Kataoka T., Takahashi N., Obata M., Horjo T.;
RT "Complete nucleotide sequence of immunoglobulin gamma2b chain gene
RL cloned from newborn mouse DNA.";
RL Nature 283:786-789(1980).
RN [2]
RP SEQUENCE FROM N.A. (MPC 11).
RX MEDLINE=80081501; Pubmed=117548;
RA Tucker P.W., Marcu K.B., Slightom J.L., Blattner F.R.;
RT "Structure of the constant and 3' untranslated regions of the murine
RL gamma 2b heavy chain messenger RNA.";
RL Science 206:1299-1303(1979).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=80081502; Pubmed=117549;
RA Tucker P.W., Marcu K.B., Newell N., Richards J., Blattner F.R.;
RT "Sequence of the cloned gene for the constant region of murine gamma
RL 2b immunoglobulin heavy chain.";
RL Science 206:1303-1306(1979).
RN [4]
RP SEQUENCE FROM N.A. (ALLELE B).
RX MEDLINE=82173203; Pubmed=6803173;
RA Olio R., Rougeon F.;
RT "Mouse immunoglobulin allotypes: post-duplication divergence of gamma
RL 2a and gamma 2b chain genes.";
RL Nature 296:761-763(1982).
RN [5]
RP CARBOHYDRATE-LINKAGE SITE THR-105.
RX MEDLINE=94216359; Pubmed=7512967;
RA Kim H., Yamaguchi Y., Masuda K., Matenaga C., Yamamoto K.,
RA Irimura T., Takahashi N., Kato K., Arata Y.;
RT "O-glycosylation in hinge region of mouse immunoglobulin G2b.";
RL J. Biol. Chem. 269:12345-12350(1994).
CC -1- SUBCELLULAR LOCATION: Secreted (potential).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Secreted;
CC IsoId=P01866-1; Sequence=Displayed;
CC Note=May be the major isoform;

```

```

CC      CC      Name=Membrane-bound;
CC      CC      IsoId=P01867-1; Sequence=External;
CC      CC      -1- PTM: O-LINKED GLYCAN CONSISTS OF GAL-GALINAC DISACCHARIDE WHICH IS
CC      CC      MODIFIED WITH 2 STATIC ACID RESIDUES.
CC      CC      -1- PTM: DISULFIDE BONDS BY HOMOLOGY WITH OTHER IG GAMMA CHAINS.
CC      CC      -1- MISCELLANEOUS: The a allele sequence is shown.
CC      CC      -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
DR      PIR; S25057; G2MS11.
DR      HSSP; P01842; 7PAB.
DR      InterPro; IPR007110; Ig-like.
DR      InterPro; IPR003597; Ig cl.
DR      InterPro; IPR003006; Ig_MHC.
DR      Pfam; PF00047; Ig; 3.
DR      SMART; SM00407; IgC1; 2.
DR      PROSITE; PS50835; IG-LIKE; 3.
DR      PROSITE; PS00290; IG_MHC; 1.
KM      Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KM      Alternative splicing; Repeat.
FT      NON_TER 1 1
FT      DOMAIN 6 98 IG-LIKE 1.
FT      DOMAIN 127 226 IG-LIKE 2.
FT      DOMAIN 235 331 IG-LIKE 3.
FT      DISULFID 15 15 INTERCHAIN (WITH A LIGHT CHAIN).
FT      DISULFID 27 82 INTERCHAIN (WITH A HEAVY CHAIN).
FT      DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
FT      DISULFID 112 112 INTERCHAIN (WITH A HEAVY CHAIN).
FT      DISULFID 115 115 INTERCHAIN (WITH A HEAVY CHAIN).
FT      DISULFID 118 118 INTERCHAIN (WITH A HEAVY CHAIN).
FT      DISULFID 150 210
FT      DISULFID 256 314
FT      CARBOHYD 105 105
FT      MOD_RES 336 336 REMOVED POST-TRANSLATIONALLY (PROBABLE).
FT      VARIANT 163 163 Q -> R (IN ALLELE B).
FT      VARIANT 194 194 T -> A (IN ALLELE B).
FT      VARIANT 300 300 M -> I (IN ALLELE B).
FT      VARIANT 301 301 L -> S (IN REF. 2 AND 3).
FT      CONFLICT 25 25 L -> P (IN REF. 2 AND 3).
FT      CONFLICT 36 36 I -> T (IN REF. 2 AND 3).
FT      CONFLICT 239 239
SQ      SEQUENCE 336 AA; 36658 MW; 7D879662607C356E CRC64;

Query Match 23.1%; Score 788.5; DB 1; Length 336;
Best Local Similarly 45.4%; Pred. No. 1,de-43;
Matches 157; Conservative 51; Mismatches 89; Indels 49; Gaps 4;

Oy      285 POLAPQVAGSGNLTALAE-----KTGKLHGEVNLVYMRATQLOKNTCEVWGPTSPKLM 339
Db      34 PESTVTWNSGSLSSSVHTPPALLQSGLYTMSSSVTPSSTWSPQVTCVAVAPAS-STT 92
Oy      340 LSLLENKEAKVSKREKPVVWLNPEAGMOCLLSDSQVLLSNIKYLPWTSTVEBPKSC 399
Db      93 VDKKLE-----PSGPISITNP-----C 109
Oy      400 DKHTPCPCAPPELLGSPVFLFPKPKDLMISRTPEVTCVVVDVSHEDPEVKFNNYVD 459
Db      110 PPKCECHKCPAPNIEGSPVFIPPNIKDMLSTLTKVTCVVVDVSEDDPDVQISWFWN 169
Oy      460 GVEVHNKAKTPREBQVSTRVYVSVLTVLHODWLNKGEKCKVSKNKAUPAPIEKTISKAK 519
Db      170 NVEVHTAQTOHREDVNSTIRVVSSTLPIQHDWMSGSEFKKVNKKDLPSEIETISKIK 229
Oy      520 GQPREPQVYTLPPSRDELTKNOVSLTCLVKGFPYSDIAVEMESNQCPENNYKTTTPVLDS 579
Db      230 GLVYAPQVYTLPPSRDELTKNOVSLTCLVKGFPYSDIAVEMESNQCPENNYKTTTPVLDS 289
Oy      580 DGSFELYSKLTVDKSRWQOGNVSQSVMEALHNNHYTKSLSPG 625
Db      290 DGSYFIYSKLNMKTSKWEKTDTSFSCNVRHGLKKNYLYLKKTISRSPG 335

RESULT 34
MUCM_MOUSE STANDARD; PRT; 476 AA.

```

[illegible]

SEQ	SEQUENCE	476 AA;	52656 MW;	804765755A204071 CMC64;
Query Match	11.0%; Score 375.5; DB 1; Length 476;			
Best Local Similarity	22.9%; Pred. No. 6.8e-17;			
Matches 127; Conservative 83; Mismatches 188; Indels 157; Gaps 18				
QY	144 ESPSSSSVOCRRSP-RGKNIQGGKTTLSVSLQELDQSGTMTCTVLQNGKVEFKIDIVL 202			
DB	3 QSPFVFPFLVSCESPLSKNLVANGCLRLDPLPSTISFTWN----- 43			
QY	203 AFQKASIVYKKEGQVFEFPFLAFTVEKLTGSGELMMQAEKASSSSKMITFDLKNKEVS 262			
DB	44 -----YQNTNEVIQ-----GIRFFPLIRGTGKYLA 68			
QY	263 VKRVYTDPKLQ-----GKKLPLHLTLPOALPOYAGSGNLTALAEKTKL 308			
DB	69 TSQVLLSPKSLILEGSDVYLVCIKHYGKRDHLVPI-----AVAE 110			
QY	309 HQEVLV-----MRATQGNLTCEWGPRTSPKMLSLKLENKAKYKREKRPVVLN 362			
DB	111 NPNVNVFVPRPDGFGPAPRKSKLICATNFTPKPIITVSWLKDGLVSGFTTDPVTIEN 170			
QY	363 ---PEA-----GMOQCLSDSQVLLESNIKVLPTWSTFVEPKSCDK 401			
DB	171 KGSITPQTKVITSLTISLIDMLNLNVYICRDHRLTLK-----NV 212			
QY	402 THTCPCPAPELLGGSVFLFPPEPKDITLMSRTPEVTGVVVDVSHEDPEVKNFYVDGV 461			
DB	213 SSTCAASPDSTIL-----FTIIPSPAD-IPLSKANLTCLVSNLNTYVE-TLNISSWASQSG 266			
QY	462 EVHNAKTPREEQNVSTRVSVLTVLHQDMLNGEKYCKYSNKLPAPIETIKSKAQ 521			
DB	267 EPLETKIKIMESHPIGTFSSAKGVASVCVEDMNNRKEPFCTVYHRDLPSPQKKFKSK--- 322			
QY	522 PRE-----POYVTLPPSRDELTK-KQVSLTCLVGFYSPDIAVEMESQO--PENNYKTT 573			
DB	333 PNEVHKHPRAVYLLPAREQNLNKRSAFTVCLVKKFSPADISVQMLQSGQLLPQEKYVTS 382			
QY	574 PPVID--SDGSFFLYSKLTVDSKRWQGNVPSCSYMBEALHNHYTQKSLSPGQLDET 631			
DB	383 APMPERGAPGFYFTHSILTVTEBENNSGETYTCVVGHEALPHLVTERTVDS-----TEG 437			
QY	632 CAEADGELDLGWT 646			
DB	438 EVNAEEEGFENLWTT 452			
RESULT 35				
EPC_MOUSE	STANDARD; PRT; 421 AA.			
AC	P06336; P01856;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	Ig epsilon chain C region.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.			
OX	NCBI_Taxid=10090;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=84236092; PubMed=6329728;			
RA	Ishida N., Ueda S., Hayashida H., Miyata T., Honjo T.;			
RT	"The nucleotide sequence of the mouse immunoglobulin epsilon gene:			
RT	comparison with the human epsilon gene sequence.";			
RL	EMBO J. 1:1117-1123(1982).			
RN	(2)			
RP	REVISIONS.			
RA	Honjo T.;			
RL	Submitted (APR-1986) to the EMBL/GenBank/DBJ databases.			
RN	(3)			
RP	SEQUENCE OF 34-421 FROM N.A.			
RX	MEDLINE=8311774; PubMed=6818553;			

```

RA Liu F.-T., Albrandt K., Sutcliffe J.G., Katz D.H.;
RT "Cloning and nucleotide sequence of mouse immunoglobulin epsilon
RL chain cDNA."
CC Proc. Natl. Acad. Sci. U.S.A. 79:7852-7856(1982).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X01857; CAA25977.1; -
DR EMBL; X01857; CAA25978.1; -
DR PIR; A02144; EHMS.
DR PIR; A02145; EHMS.
DR HSSD; P01854; IIGF.
DR InterPro; IPR007110; Ig-1like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig1; 4.
DR SMART; SM00407; IGc1; 2.
DR PROSITE; PS00835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 3.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein.
FT NON TER 1 1
FT DOMAIN 1 90 CH1.
FT 198 197 CH2.
FT DOMAIN 198 304 CH3.
FT 305 421 CH4.
FT DISULFID 23 75 BY SIMILARITY.
FT DISULFID 121 180 BY SIMILARITY.
FT DISULFID 226 285 BY SIMILARITY.
FT DISULFID 330 392 BY SIMILARITY.
FT CARBOHYD 43 43 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 421 AA; 47320 MW; 8F909E1F30A06B47 CRC64;
-----
Query Match 10 7%; Score 354.5; DB 1; Length 421;
Best Local Similarity 28.7%; Pred. No. 2,9e-16;
Matches 123; Conservative 75; Mismatches 163; Indels 67; Gaps 21;
-----
QY 223 FPLAFTVETKLNGSGELMWAQERASSKSMITFDLKNKE--VSVKRVTDPPKLGQMKLPL 280
DB 29 FPNPTVTV-----WYSD--SLNMSTVNFPAISGLKTKTSQVYTWGR--SAKNPTC 75
QY 281 HLTTPQALPYAGSGNLTL---ALEAKTGKIHQEVNLVWKAATQLOKNTLCEVWGPTSP 336
DB 76 HVTNH---PSFNESRTILVRPNVITEPTLELHSCDPAHSHST---IQLYCFIYHILN 129
QY 337 KLMISLKLENKAKYKREKRPVWLVNPEAGMQLCLISGQV--LLESIXKVLPTMSTFVE 395
DB 130 DVASVSMWMDREITDTLAQT---VLIKEGK--LASTCSKNTTEQQMSESTFTCVT 183
QY 396 PKSCDKTTCPPCPAPPELLGSPVLPFPKPKDTLMIRTEPVTCVVVDV--SHEDPEVKF 454
DB 184 SQGVYDYLHTRCRPHF--PRGVITYLIRPSPLD--LYNGAKRKLTLVVDLSEKKNVNTW 241
QY 455 N-----WYVDGVEVHNAAKTKPREEOYNSTYRVSVLYTLVHODLMNGKEYCKRVS 503
DB 242 NOEKKTSSASQWY---TKHNH-----NATTSITSLIPVAKWIMGYGQICVD 288
QY 504 NKALPAPLEKTIISRAKKGPREBOVYTLPPSRDELTKNOVSLTCLVKGYPYSDIAVEMESN 563
DB 289 HPDFPKPIVRISITKTPGGRSAPEVVPFPPEEB--SEDKRTLTCLIONFFPEDISVOMLGD 347

```

```

QY 564 GQPENN--YKTPPEVLDSGS---FPLYSKLTVDKSRMOGNVFSQSVMEALHN-HYTD 617
DB 348 GKLISSNGHSTTTP--LNSNSNGGFFIFSRLEVAKTLMTQRKQTCOVHIALQKPRLE 406
QY 618 KSLSLSPG 625
DB 407 KTIISTSLG 414
-----
RESULT 36
MOC_MOUSE STANDARD; PRT; 455 AA.
ID MOC_MOUSE
AC P01872;
DC 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig mu chain C region secreted form.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81076590; PubMed=6255422;
RA Kawakami T., Takahashi N., Honjo T.;
RT "Complete nucleotide sequence of mouse immunoglobulin mu gene and
RL Nucleic Acids Res. 8:3933-3945(1980).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=82051295; PubMed=6795090;
RA Goldberg G.I., Vanin R.F., Zolka A.M., Blattner F.R.;
RT "Sequence of the gene for the constant region of the mu chain of
RL Balb/c mouse immunoglobulin.",
RN [3]
RP SEQUENCE FROM N.A. (MYELOMA TEPCL83).
RX MEDLINE=81165562; PubMed=6260591;
RA Auftray C., Rougeon F.;
RT "Nucleotide sequence of a cloned cDNA corresponding to secreted mu
RL chain of mouse immunoglobulin.",
RN [4]
RP SEQUENCE (MYELOMA PROTEIN MOPC 104E).
RX MEDLINE=79223904; PubMed=111247;
RA Kehry M.R., Sibley C.H., Fuhrman J.S., Schilling J.W., Hood L.B.;
RT "Amino acid sequence of a mouse immunoglobulin mu chain.",
RL Proc. Natl. Acad. Sci. U.S.A. 76:2932-2936(1979).
RN [5]
RP REVISION (MOPC 104E).
RX MEDLINE=83075344; PubMed=6816276;
RA Kehry M.R., Fuhrman J.S., Schilling J.W., Rogers J., Sibley C.H.,
RT Hood L.B.;
RT "Complete amino acid sequence of a mouse mu chain: homology among
RL heavy chain constant region domains.",
RL Biochemistry 21:5415-5424(1982).
CC -1- SUBCELLULAR LOCATION: Secreted (Probable).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=During differentiation, B lymphocytes switch from
CC expression of isoform Membrane-bound to isoform Secreted;
CC Name=Secreted;
CC IsoId=P01872-1; Sequence=Displayed;
CC Name=Membrane-bound;
CC IsoId=P01873-1; Sequence=External;
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@slb-sib.ch).
CC -----
CC EMBL; V00827; -; NOT_ANNOTATED_CDS.
CC PIR; A02166; MHMS.
CC HSP; P01857; 1FC1.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003597; Ig_c1.
CC InterPro; IPR003006; Ig_MHC.
CC Pfam; PF00047; Ig; 4.
CC SMART; SM00407; IgC1; 2.
CC PROSITE; PSS00230; IG_MHC; 3.
CC PROSITE; PSS00230; IG_LIKE; 4.
CC Immunoglobulin domain; IG_MHC; 3.
CC Immunoglobulin C region; Glycoprotein;
KW Alternative splicing.
FT NON_TER 1
FT DOMAIN 1 105 CH1.
FT DOMAIN 106 217 CH2.
FT DOMAIN 218 324 CH3.
FT DOMAIN 325 436 CH4.
FT DOMAIN 437 455 C-TERMINAL SEGMENT OF SECRETED FORM.
FT DISULFID 14 14 INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
FT DISULFID 28 89 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 136 199 BY SIMILARITY.
FT DISULFID 216 216 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 246 305 BY SIMILARITY.
FT DISULFID 293 293 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 353 415 BY SIMILARITY.
FT DISULFID 454 454 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .).
FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .).
FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .).
FT CARBOHYD 281 281 N-LINKED (GLCNAC. . .).
FT CARBOHYD 442 442 S -> N (IN MOPC 104E).
FT CARBOHYD 78 78 S -> Q (IN MOPC 104E).
FT VARIANT 101 101 T -> T (IN MOPC 104E).
FT VARIANT 226 226 N -> T (IN MOPC 104E).
FT VARIANT 258 258 N -> S (IN TEPC183).
FT VARIANT 258 258 L -> K (IN TEPC183 AND MOPC 104E).
FT VARIANT 368 368 L -> K (IN TEPC183 AND MOPC 104E).
FT VARIANT 455 455 4CB557CB602F9851 CRC64;
SQ SEQUENCE 455 AA; 50101 MW; 4CB557CB602F9851 CRC64;

Query Match 10.7%; Score 364; DB 1; Length 455;
Best Local Similarity 22.8%; Pred. No. 3.5e-16;
Matches 122; Conservative 80; Mismatches 180; Indels 152; Gaps 17;

QY 144 ESPRSSPSVQCSRP-RGKRIQGGKTLVSQLELDSSGWTCTVLONOKKVEKIDIVVL 202
DB 3 QSEPNVFPVSCSPSLSDKVLAMGCLARDLPSTISFTWN----- 43
QY 203 AFQKASSIVYKKEGQVFEFSFLAFTVEKLTGSGELMQAERASSKSWITFDLKNKEVS 262
DB 44 -----YQNTTEVIO-----GRTPTLTGSKYLA 68
QY 263 VKRVTQDPKLQW-----GKPLPLHLTLPLALPOVAGSGNLTLEAKTGKL 308
DB 69 TSOVLVLSKSIIEGSDVYLCKIHYGGKRDHLVLP-----AVAMM 110
QY 309 HOEVNLV-----MATOLQKULTEWGPSPKMLSLKLENKAKYSKKEKPYWVLN 362
DB 111 NPVNVFVPDRGFSGPAPRPSKSLICEATNFTPKPTVSWLMDKGLVESGFTTDPVTJIN 170
QY 363 -----PEA-----GWMQCLSDSGQVLLESNIKVLPTMSTVEPRPSCK 401
DB 171 KSTPTTYAVISTLTSEIDMLANLYTRKVRHRLGTLFK-----NV 212
QY 402 THTCPEPAPELLGDSVFLFPKPKDLMISKTPEVTCVVVDVSHEDPEVFNMYVDGV 461
DB 213 SSTCAASPSTDL-----FTIPSPAD-IFLSKANLTLGVSNLAYE-TLNISSMVSOG 266
QY 462 EVHNAATKREBEVNSTYVSLYTLADMLNGKSKYKKVSKNKAIPAEITKTSIAKQO 521
DB 267 EPLETKIKIMESHPIGNTGSAKVASVCVBDMMNRKEFVCTVTHRDLSPQKFKISK--- 322

```

```

QY 522 PRE-----POVYTLPPSRDELTKNOVSLTCLVKGFYPSDIAVEMESNQ--PENNYKTT 573
DB 323 PNEVHKHPAVYLLPAPRQNLRESATVTLGVKFSRPAIDIVQVLQSGQLLPQEKYTS 382
QY 574 PVLVD--SDGSFPLYSKLTVDKSRNOQGVNFCSSVMEALHNHYTKSISLSPG 625
DB 383 APMPERGAGPFGFYTHSILVTEENWNSGETYTCVGVHEALPHLVTERVTDKSTG 436

RESULT 37
MUC_HUMAN
ID MUC_HUMAN STANDARD; PRT; 454 AA.
AC P01871;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 19 mu chain C region.
GN IGHM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OC NCBI_Taxid=9606;
RN [1]
RP SEQUENCE OF 1-434 FROM N.A.
RX MEDLINE=90332450; PubMed=2115996;
RA Friedlander R.M.; Nussenzweig M.C.; Leder P.;
RT "Complete nucleotide sequence of the membrane form of the human Igm
RT heavy chain."
RL Nucleic Acids Res. 18:4278-4278 (1990).
RN [2]
RP SEQUENCE (WALDENSTROM'S MACROGLOBULIN GAL).
RX MEDLINE=75059123; PubMed=4803843;
RA Watanabe S.; Barnikol H.U.; Horn J.; Bettram J.; Hilschmann N.;
RT "The primary structure of a monoclonal Igm-immunoglobulin
RT (macroglobulin Gal.); II: the amino acid sequence of the H-chain (mu-
RT type), subgroup H III. Architecture of the complete Igm-molecule."
RL Hoppe-Seyler's Z. Physiol. Chem. 354:1505-1509 (1973).
RN [3]
RP REVISIONS (GAL).
RX MEDLINE=81066716; PubMed=6777162;
RA Mihaesco E.; Barnikol-Watanabe S.; Barnikol H.U.; Mihaesco C.;
RA Hilschmann N.;
RT "The primary structure of the constant part of mu-chain-disease
RT protein BOR."
RL Eur. J. Biochem. 111:275-286 (1980).
RN [4]
RP SEQUENCE (WALDENSTROM'S OU), DISULFIDE BONDS, AND CARBOHYDRATES.
RX MEDLINE=74005511; PubMed=4742735;
RA Putnam F.W.; Florent G.; Paul C.; Shinoda T.; Shimizu A.;
RT "Complete amino acid sequence of the mu heavy chain of a human Igm
RT immunoglobulin."
RL Science 182:287-291 (1973).
RN [5]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=82059479; PubMed=6795593;
RA Rabbits T.H.; Forster A.; Milstien C.P.;
RT "Human immunoglobulin heavy chain genes: evolutionary comparisons of
RT C mu, C delta and C gamma genes and associated switch sequences."
RL Nucleic Acids Res. 9:4509-4524 (1981).
RN [6]
RP SEQUENCE OF 299-387 AND 438-454 FROM N.A.
RX MEDLINE=81077306; PubMed=6777778;
RA Dolby T.W.; Devunuo J.; Croce C.M.;
RT "Cloning and partial nucleotide sequence of human immunoglobulin mu
RT chain cDNA from B cells and mouse-human hybridomas."
RL Proc. Natl. Acad. Sci. U.S.A. 77:6027-6031 (1980).
CC -I- MISCELLANEOUS: All 4 combinations of the S/G and V/G polymorphisms
CC at positions 192 and 216 have been observed in human mu chains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X17115; CAA34971.1; ALT_SEQ.
 DR EMBL: X57086; -, NOT_ANNOTATED_CDS.
 DR HSSP: P01857; 1FCL.
 DR Genew: HGNC:5541; IGHM.
 DR MIM: 147020; -.
 DR GlycoSiteDB: P01871; -.
 DR GO: GO:0005624; C:membrane fraction; NAS.
 DR GO: GO:0003823; F:antigen binding; NAS.
 DR GO: GO:0006955; P:immune response; NAS.
 DR InterPro: IPR007110; IG-like.
 DR InterPro: IPR003597; IG cl.
 DR InterPro: IPR003006; IG_MHC.
 DR Pfam: PF00047; Ig; 4.
 DR SMART: SM00407; IGcl; 3.
 DR PROSITE: PS50835; IG_LIKE; 4.
 DR PROSITE: PS00290; IG_MHC; 3.
 KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
 KW Polymorphism.

FT NON_TER 1 1
 FT DOMAIN 1 105 CH1.
 FT DOMAIN 106 218 CH2.
 FT DOMAIN 219 324 CH3.
 FT DOMAIN 325 454 CH4.
 FT DISULFID 14 14 INTERCHAIN (WITH LIGHT CHAIN).
 FT DISULFID 28 88
 FT DISULFID 135 198
 FT DISULFID 215 215 INTERCHAIN (WITH HEAVY CHAIN).
 FT DISULFID 245 304
 FT DISULFID 292 292 INTERCHAIN (WITH HEAVY CHAIN IN ANOTHER
 SUBUNIT).
 FT DISULFID 352 414
 FT DISULFID 453 453 INTERCHAIN (WITH HEAVY CHAIN).
 FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 210 210 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 273 273 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 441 441 N-LINKED (GLCNAC. . .).
 FT CARBOHYD /FTID=CAR_000219.
 FT VARIANT 192 192 S->G.
 FT VARIANT /FTID=VAR_003903.
 FT VARIANT 216 216 V->G (in dbSNP:12365).
 FT VARIANT /FTID=VAR_003904.
 FT VARIANT /FTID=VAR_003904.
 SQ SEQUENCE 454 AA; 49556 MM; 21EC72BADC56922E CRC64;

Query Match 10.5%; Score 359.5; DB 1; Length 454;
 Best Local Similarity 27.6%; Pred. No. 6.8e-16;
 Matches 83; Conservative 62; Mismatches 111; Indels 45; Gaps 10;

QY 332 GPTSKMLSLKLNKKEKAVSKREKPVWVLYNPEAGMOCCLSDSCQVLE-SNIVVLTW 390
 DB 173 GPTTKVSTLTIKESD-----M-----SQSMFCRDVHRLTFOQNASMNCVDPQ 219
 QY 391 SPPVPKSCDKTHTCPCPAPPELLGSPVFLFPPPKDTLMSRTPEYTCVVDVSHSDP 450
 DB 220 DTAIR-----VFALPPS-FASIFLTSTKTLCLVLTDTITTD- 254
 QY 451 EVKFNWYGVGVHNAKTKPREQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAP 510
 DB 255 SVTISWTHNGDAVVAHTHTNISHSNAPSAVGAISCEDDWNSGERFTCTVHTDLPSP 314
 QY 511 IETKTSKAKGP-REPOVYTLTPSRDELTKQVSLTCLVYKGFYSDDIAVENESNGGP-- 566
 DB 315 LKQITSRPGVALHHPDYVLLPAPAEQLMLRESAITTLCLVYGFSPADVFVQVMORGQPLS 374
 QY 567 ENNYTTPPVLD--SDGSFFLYSKLTVDKSRWQGNVFSQVMHEALHNHYTKSLSP 624
 DB 375 PEKYTSAWPEPQAPGRYFAHSILTVEEWNIGETTCVVAHALLPRVTERIVDKST 434
 QY 625 G 625

DB 435 G 435

RESULT 38
 EPC_HUMAN
 ID EPC_HUMAN STANDARD; PRT; 428 AA.
 AC P01854;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Ig epsilon chain C region.
 GN IGH.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8316897; PubMed=6300763;
 RA Seno M., Kurokawa T., Ono Y., Onda H., Sasada R., Igarashi K.,
 RA Kikuchi M., Sugino Y., Nishida Y., Honjo T.;
 RT "Molecular cloning and nucleotide sequencing of human immunoglobulin
 epsilon chain cDNA.";
 RL Nucleic Acids Res. 11:719-726(1983).
 RN [2]
 RP SEQUENCE FROM N.A. AND VARIANT LEU-359.
 RX MEDLINE=83001945; PubMed=6288268;
 RA Max E.E., Battey J., Ney R., Kirsch I.R., Leder P.;
 RT "Duplication and deletion in the human immunoglobulin epsilon genes.";
 RL Cell 29:691-699(1982).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84236029; PubMed=6234164;
 RA Flanagan J.G., Rabbits T.H.;
 RT "The sequence of a human immunoglobulin epsilon heavy chain constant
 region gene, and evidence for three non-allelic genes.";
 RL EMBO J. 1:655-660(1982).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84207910; PubMed=6327276;
 RA Ueda S., Nakai S., Nishida Y., Hisajima H., Honjo T.;
 RT "Long terminal repeat-like elements flank a human immunoglobulin
 epsilon pseudogene that lacks introns.";
 RL EMBO J. 1:1539-1544(1982).
 RN [5]
 RP PRELIMINARY SEQUENCE (MYELOMA PROTEIN ND).
 RA Benrich H.H., Johansson S.G.O., von Bahr-Lindstrom H.;
 RL (in) Bach M.K. (eds.);
 RT Immediate hypersensitivity: modern concepts and developments, pp.1-36,
 RL Marcel Dekker, New York (1978).
 RN [6]
 RP SEQUENCE OF 1-40; 68-114 AND 427-428 FROM N.A.
 RX MEDLINE=83065234; PubMed=6815656;
 RA Kenten J.H., Moigaard H.V., Houghton M., Derbyshire R.B., Viney J.,
 RA Bell L.O., Gould H.J.;
 RT "Cloning and sequence determination of the gene for the human
 immunoglobulin epsilon chain expressed in a myeloma cell line.";
 RL Proc. Natl. Acad. Sci. U.S.A. 79:6661-6665(1982).
 RN [7]
 RP 3D-STRUCTURE MODELING.
 RX MEDLINE=87089848; PubMed=3796618;
 RA Padlan E.A., Davies D.R.;
 RT "A model of the FC of immunoglobulin E.";
 RL Mol. Immunol. 23:1063-1075(1986).
 CC -I- SIMILARITY: Contains 4 immunoglobulin-like domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC	or send an email to license@lsb-sib.ch).	
CC	-----	
DR	EMBL; L00022; AAB59424.1; ALT_INIT.	
DR	PIR; A22771; EHHU.	
DR	PDB; 1IGE; 15-JUL-92.	
DR	PDB; 1FP5; 30-JAN-02.	
DR	PDB; 1G84; 16-MAY-01.	
DR	PDB; 100V; 18-SEP-02.	
DR	GeneW; HGNC:5522; IGHE.	
DR	MIM; 147180; -.	
DR	GO; GO:0003823; P:antigen binding; NAS.	
DR	GO; GO:0006955; P:immune response; NAS.	
DR	InterPro; IPR007110; Ig-1ike.	
DR	InterPro; IPR003587; Ig-cl.	
DR	InterPro; IPR003006; Ig_MHC.	
DR	Pfam; PF00047; Igc1; 4.	
DR	SMART; SM00407; Igc1; 4.	
DR	PROSITE; PS50835; IG_LIKE; 4.	
DR	PROSITE; PS00290; IG_MHC; 3.	
KM	Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;	
KW	3D-structure; Repeat.	
FT	NON_TER	1
FT	DOMAIN	6
FT	DOMAIN	112
FT	DOMAIN	210
FT	DOMAIN	214
FT	DOMAIN	318
FT	DOMAIN	324
FT	DISULFID	423
FT	DISULFID	14
FT	DISULFID	14
FT	DISULFID	15
FT	DISULFID	105
FT	DISULFID	29
FT	DISULFID	85
FT	DISULFID	121
FT	DISULFID	121
FT	DISULFID	193
FT	DISULFID	209
FT	DISULFID	239
FT	DISULFID	299
FT	DISULFID	345
FT	CARBOHYD	405
FT	CARBOHYD	21
FT	CARBOHYD	21
FT	CARBOHYD	48
FT	CARBOHYD	48
FT	CARBOHYD	99
FT	CARBOHYD	99
FT	CARBOHYD	146
FT	CARBOHYD	146
FT	CARBOHYD	252
FT	CARBOHYD	252
FT	CARBOHYD	275
FT	CARBOHYD	275
FT	VARIANT	359
FT		359
FT	STRAND	110
FT	STRAND	113
FT	STRAND	117
FT	HELIX	122
FT	HELIX	124
FT	TURN	125
FT	TURN	125
FT	STRAND	130
FT	STRAND	141
FT	STRAND	141
FT	STRAND	146
FT	STRAND	155
FT	STRAND	155
FT	STRAND	159
FT	STRAND	161
FT	STRAND	165
FT	STRAND	166
FT	TURN	169
FT	STRAND	172
FT	STRAND	181
FT	HELIX	182
FT	HELIX	186
FT	TURN	187
FT	TURN	188
FT	STRAND	192
FT	STRAND	196
FT	TURN	198
FT	TURN	199
FT	TURN	201
FT	TURN	202
FT	STRAND	205
FT	STRAND	207
FT	STRAND	218
FT	TURN	222
FT	TURN	222
FT	TURN	226
FT	TURN	229
FT	TURN	231
FT	TURN	232
FT	STRAND	236
FT	STRAND	242
FT	STRAND	252
FT	STRAND	252
FT	STRAND	255
FT	TURN	257
FT	TURN	258
FT	TURN	259
FT	STRAND	260
FT	STRAND	266
FT	TURN	265
FT	STRAND	271
FT	STRAND	274
FT	TURN	275
FT	STRAND	276
FT	HELIX	277
FT	HELIX	285
FT	HELIX	288
FT	TURN	293
FT	TURN	294
FT	TURN	294

[illegible]

```

RA Bernstein K.E., Alexander C.B., Reddy E.P., Mage R.G.
RT "Complete sequence of a cloned cDNA encoding rabbit secreted mu-chain
RL J. Immunol. 132:490-495(1984).
CC -I- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=During differentiation, B lymphocytes switch from
CC Name=Membrane-bound;
CC IsoId=P04221-1; Sequence=Displayed;
CC Name=Secreted;
CC IsoId=P03986-1; Sequence=External;
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K01357; AAA1293.1; -
DR PIR; A02165; MHRM.
DR HSP; P01842; 7FAB.
DR InterPro; IPR007110; Ig_1-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_4.
DR SMART; SM00407; IgC1.2.
DR PROSITE; PS50835; Ig_LIKE; 4.
DR PROSITE; PS00290; Ig_MHC; 3.
DR Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Alternative splicing; Transmembrane.
FT NON_TER 1 1
FT DOMAIN 1 106 CH1.
FT DOMAIN 107 222 CH2.
FT DOMAIN 223 327 CH3.
FT TRANSMEM 328 458 CH4.
FT DISULFID 459 476 POTENTIAL.
FT DISULFID 14 14 INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
FT DISULFID 28 90 BY SIMILARITY.
FT DISULFID 137 200 BY SIMILARITY.
FT DISULFID 219 219 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 249 308 BY SIMILARITY.
FT DISULFID 296 308 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 356 418 BY SIMILARITY.
FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 212 212 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 284 284 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 479 AA, 52351 MW, 689C637A47BE19FC CRC64;

Query Match 10.5%; Score 358.5; DB 1; Length 479;
Best Local Similarity 24.6%; Pred. No. 8.4e-16;
Matches 121; Conservative 83; Mismatches 175; Indels 113; Gaps 21;

```

```

OY 406 PCPAPILLGSPVFLPPPKDTLMISRTPEVTCVVDVSHEDPEVKFNWYDGEVHN 465
DB 220 STTSP-----GIQFPIAPSPADI-FLSKGARILCLVTDLTYY-SLINS-----ASHN 268
OY 466 AKT-----KPREQYNSTRVAVSVLVYVHODWLNKGEYKCKVSKALPAPIEKTISKAG 520
DB 269 GKALDTMNITESHSPNTFSGMGEASVCADMEGSEQFTCTVTHADPPFLKHTISKRE 328
OY 521 QPRP-PQVYTLPPSRDLT-KNQVSLTCLVKGFPSPDIAVWESNGQP--ENNKTTPPV 576
DB 329 VAKHPVAVVLPAPREQLVRESATVTCLVKGSFPADVQWQQRGQPLSSDKVTAPAPA 388
OY 577 LD--SDSPFLYSKLVYDKSRMOQGVFSSVMEALHNHYTKSLSPQLQDETCAE 634
DB 389 PEPAPEGLYTHSLTYLTYEDDMNSGEFTCVVGHFALPHWTERTVKS-----TEGEVG 443
OY 635 AQDGEDLGLMTT 646
DB 444 AEEGPFENLWTT 455

RESULT 40
EPC_RAT ID EPC_RAT STANDARD; PRT; 429 AA.
AC P01855;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig epsilon chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A. (IMMUNOCYTOMA IR2).
RC STRAIN=LOU/C/MSL;
RX MEDLINE=83064537; PubMed=6292865;
RA Hellman L., Petersson U., Engstrom A., Karlsson T., Bennich H.;
RT "Structure and evolution of the heavy chain from rat immunoglobulin
RT E.";
RL Nucleic Acids Res. 10:6041-6049(1982).
RN [2]
RP SEQUENCE OF 168-342 FROM N.A. (MYELOMA IR162).
RX MEDLINE=83182019; PubMed=6820340;
RA Kindsvogel W.R., Reddy E.P., Moore J.M., Fauser C.H. Jr.;
RT "A cloned cDNA probe for rat immunoglobulin epsilon heavy chain:
RT construction, identification, and DNA sequence.";
RL DNA 1:335-343(1982).
RN [3]
RP SEQUENCE OF 205-306 FROM N.A.
RX MEDLINE=82174576; PubMed=680238;
RA Hellman L., Petersson U., Bennich H.;
RT "Characterization and molecular cloning of the mRNA for the heavy
RT (epsilon) chain of rat immunoglobulin E.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:1264-1268(1982).
CC -I- SIMILARITY: Contains 4 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J00744; AAA41379.1; ALT_INIT.
DR PIR; A93442; EHRT.
DR HSP; P01854; 1IGE.
DR InterPro; IPR007110; Ig_1-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_4.

```

```
DR SMART, SM00407, IGc1, 1.
DR PROSITE, PS50835, IG_LIKE, 4.
DR PROSITE, PS00290, IG_MHC, 3.
KW Immunoglobulin domain, Immunoglobulin C region, Repeat.
FT NON_TER 1
FT DOMAIN 1
FT DOMAIN 6 89 IG-LIKE 1.
FT DOMAIN 103 201 IG-LIKE 2.
FT DOMAIN 205 305 IG-LIKE 3.
FT DOMAIN 314 414 IG-LIKE 4.
FT CONFLICT 168 168 R -> N (IN REF. 2).
FT CONFLICT 308 308 P -> L (IN REF. 2).
SQ SEQUENCE 429 AA; 48671 MW; D2970B34EFA7B0 CRC64;

Query Match
Best Local Similarity 27.8%; Score 355.5; DB 1; Length 429;
Matches 101; Conservative 69; Mismatches 114; Indels 79; Gaps 16;

QY 324 KNLTCEWGPSTPEKMLSLKENKAKVSKREKPVVVLNPEAGM----- 367
DB 74 KNFTCHV--THASTFVS-----DLTIRAPVMTKPTVDLHSSCDPNAFHTIOL 123
QY 368 -----WQCLSD-----SGVYLESNITKVLPTWS---TPVEPKSCDKT 402
DB 124 YCFVYGHIONDYSIH--LMDDRKIYETHAQNVLKEEGKLASTYSRLNITQQQWMSBST 181
QY 403 HICRP-----PCPAPRLGGPSVLEFPKPKDTLMSRTEPVTCVVVDVSHED 449
DB 182 FTCKKTSQGBNVAHTRRCSDD--PRGVITTYLIPSPPLD--LVENGPKTLCLVLDSESE 239
QY 450 PEVKEFMYVDGVENHAKTRPREEQNSTYRVVSVTLVHODVLNKEYCKVSNKALPA 509
DB 240 -NITVWVERKKSSISASQSRSTKHNNATTSITSLPVAKAKMIEEGYQCRVDHHPFK 298
QY 510 PLEKTSKAKGQPREPVYT--LPSPRDELTKQVSLTCLVKGFYPSDIAVEW--ESNGQP 566
DB 299 PIYRSITKAPGKSAEYVFLPPEEEB--KDKRTLTCLIQNFPEPDISVQWLQDSKLIIP 356
QY 567 ENNYKTPPYLSDSGS---FELYSKLTVDKSRWQGNVSCSMHEALIN--HYTQKSLSL 622
DB 357 KQHSHTTTP-LKYNQSNQRFIFSRLEVTALMTQKPTCRVTHLEALPRKLEKTIISK 415
QY 623 SPG 625
DB 416 SLG 418

RESULT 41
MUC_MESAU
ID MUC_MESAU STANDARD; PRT; 454 AA.
AC P06337;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE 15 mu chain C region.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,
OC Mammalia, Eutheria, Rodentia, Sciurognathi, Muridae, Cricetinae;
OC Mesocricetus.
OC NCBI_TaxID=10036;
OX [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=85297761; PubMed=2994005;
RA McGuire K.L., Duncan W.R., Tucker P.W.;
RT "Phylogenetic conservation of immunoglobulin heavy chains: direct
RT comparison of hamster and mouse Cmu genes.";
RL Nucleic Acids Res. 13:5611-5628(1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage is for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, X02804, CAA26574.1; -.
DR HSSP, F01854, 1IGB.
DR InterPro, IPR007110, IG-1like.
DR InterPro, IPR003597, IG_c1.
DR InterPro, IPR003006, IG_MHC.
DR Pfam, PF00047, igf, 4.
DR SMART, SM00407, IGc1, 2.
DR PROSITE, PS50835, IG_LIKE, 4.
DR PROSITE, PS00290, IG_MHC, 3.
KW Immunoglobulin domain, Immunoglobulin C region, Glycoprotein.
FT NON_TER 1
FT DOMAIN 1
FT DOMAIN 106 218 CH1.
FT DOMAIN 219 324 CH2.
FT DOMAIN 325 454 CH3.
FT DOMAIN 325 454 CH4.
FT DISULFID 13 13 INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
FT DISULFID 27 88 BY SIMILARITY.
FT DISULFID 135 198 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 215 215 BY SIMILARITY.
FT DISULFID 245 304 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 292 292 BY SIMILARITY.
FT DISULFID 352 414 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 453 453 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 210 210 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 441 441 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 454 AA; 50196 MW; 4DA2134612BE1469 CRC64;

Query Match
Best Local Similarity 26.0%; Score 355.5; DB 1; Length 454;
Matches 120; Conservative 82; Mismatches 167; Indels 93; Gaps 19;

QY 223 FPLA-----FTVEKLTSGGELMWOERASSKSWITFDLKNKEVSKRVTDQPKLQWGG 277
DB 8 FPLVCSBPPLSDENLVAMGCLARDPLPSISFSW---NYQNSKSVNGQVRFPTLRMGSK 64
QY 278 -----LPLHLITPOLPYAGSGNLTALAEATGTLHGEVNLVNRATOLQCN----- 325
DB 65 YAATSQVFLP-----PKSVLE--GSDXY-LVCKVHAGNTNDRVPIVVTENMNVSVF 116
QY 326 -----LICEVWGPSTPEKMLSLKENKAKVSKREKPVVVLNPEAGMQ 369
DB 117 VESRDAPSGPAPRKSRRLFCESNSFPKQITVSMLRDGRVXSGFTTEFV---TPB----- 168
QY 370 CLLSDSGQ-----VLLESNIKVLPTWSTPEVEKSC---DKTHGCPAPRL 414
DB 169 ----DRGSGPRTYKIVISLTITTESDMLNLSYTCVDRHGLTFMKNVSTCAASPTDIO 224
QY 415 GGPSTFLPPPKKDT--LMSRTEPVTCVVVDVSHEDPEVKRMYVDGVENHAKTRPRE 473
DB 225 A-----FPPEPSFVGIFLNKSAITLCVTMLATYD--TLNLSWSSRSSEPEETKTLSES 277
QY 474 QYNSTYRVSVTLVHODVLNKEYCKVSNKALPAPIEKTISKAKGQPRE-----POVY 528
DB 278 HENGTFSAIGEANVCVEWDSKKEVCTVTHRDLPSPQKTIISK-----PRENNKTPPAVY 333
QY 529 TLPSRDEL--TKNQVSLTCLVKGFYPSDIAVEWESNGQP--ENNYKTPPYLSDSGS--F 583
DB 334 QQPLAREQDLIRESATVCLVVGFSPPADIPQWLRQGPLQDKVTSAPMREBPQAPHLV 393
QY 584 FLYSKLTVDKSRWQGNVPSGVNHEALHNHYTQKSLSLSPG 625
DB 394 FTHSVLTVTEEWNSGETTYTCVVGHEALPMVTEKTVDRSTG 435

RESULT 42
```

```

MUCB_HUMAN          STANDARD;          PRT;          391 AA.
ID   MUCB_HUMAN      20-MAR-1987 (Rel. 04, Created)
AC   P04220;
DT   20-MAR-1987 (Rel. 04, Last sequence update)
DE   Ig mu heavy chain disease protein (BOR).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=84184186; PubMed=6425189;
RA   Barnikol-Watanabe S., Mihaesco E., Mihaesco C., Barnikol H.U.,
RA   Hilschmann N.;
RT   "The primary structure of mu-chain-disease protein BOR. Peculiar
RT   amino-acid sequence of the N-terminal 42 positions.";
RL   Hoppe-Seyler's Z. Physiol. Chem. 365:105-118(1984).
CC   -1- MISCELLANEOUS: This protein has no V region homology or CH1
CC   region.
DR   PIR: A02163; MHUBT.
DR   HSBP: P01857; 1FC1.
DR   GO: GO:0005624; C:membrane fraction; NAS.
DR   GO: GO:0003823; F:antigen binding; NAS.
DR   GO: GO:0006955; P:immune response; NAS.
DR   InterPro: IPR007110; IG_1like.
DR   InterPro: IPR003597; IG_CL.
DR   InterPro: IPR003006; IG_MHC.
DR   Pfam: PF00047; Ig_3.
DR   SMART: SM00407; IgC1; 2.
DR   PROSITE: PS00835; IG_LIKE; 3.
DR   PROSITE: PS00290; IG_MHC; 3.
KW   Immunoglobulin domain; Immunoglobulin C region.
FT   DOMAIN          1      42      PRE-C-PART (NO V REGION HOMOLOGY).
FT   DOMAIN          43     155     CH2.
FT   DOMAIN          156     261     CH3.
FT   DOMAIN          262     391     CH4.
SQ   SEQUENCE          391 AA; 43057 MW; 9100843AF0CF021A CRC64;

Query Match          10.4%; Score 353.5; DB 1; Length 391;
Best Local Similarity 27.8%; Pred. No. 1.3e-15;
Matches 84; Conservative 61; Mismatches 110; Indels 47; Gaps 11;

OY   332 GPTSPKMLSLKLEKKAQVSKREKPVVNLNPEAGMOCLSDSGOVLLESNIKVL--PT 389
DB   110 GPTTYKVTSLTIKESD-----WL---GQSMFTGCRVDHG-LTFQGNASMGCPD 155
OY   390 WSTPVEPKSCDKHTCCPCAPBELIGSPVFLFPKPKDTLMISRTPEVTCVVVDVSHED 449
DB   156 QDTAIR-----VFALPPS--PASIFLTSTCLTCLVDTLTYD 191
OY   450 PEVKNNVVVDGVNNAKTPREBOYNSTYRVSVLTVLHODMNGKVKYKKNKALPA 509
DB   192 -SVTISMRQGEAVKHTNISHPNATFSAVGEASICEPDWDSGEFTCTVHTDIPS 250
OY   510 PIEKTISSAKGP-REPQVYTLPSRDELTKNQVSLTCLLVGFPSDIAVEMESNGP- 566
DB   251 PLKQTISSPKGVALLHRPVVILLPRAREGLNLRBSATITCLVTGSPAPVFGVMQGRGPL 310
OY   567 -ENNYKTPPVLD--SDGSFPLYSKLTVDKSRMOQGNVFGCSVMEALHHNYTKSLSL 623
DB   311 SPEKVTISAPWEPQAPGRVFAHSITLVSEEMWTGETYTCVVAHEALPNVTERTVDS 370
OY   624 PG 625
DB   371 TG 372

```

```

DT   23-OCT-1986 (Rel. 02, Created)
DT   23-OCT-1986 (Rel. 02, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Ig mu chain C region secreted form.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A. (A2 ALLOTYPE).
RX   MEDLINE=84088930; PubMed=6418803;
RA   Bernstein K.E., Alexander C.B., Reddy E.P., Mage R.G.;
RT   "Complete sequence of a cloned cDNA encoding rabbit secreted mu-chain
RT   of Vha2 allotype: comparisons with Vha1 and membrane mu sequences.";
RL   J. Immunol. 132:490-495(1984).
CC   -1- SUBCELLULAR LOCATION: Secreted (Probable).
CC   -1- ALTERNATIVE PRODUCTS:
CC   Event=Alternative splicing; Named isoforms=2;
CC   Comment=During differentiation, B lymphocytes switch from
CC   expression of isoform Membrane-bound to isoform Secreted;
CC   Name=Secreted;
CC   IsoId=P03988-1; Sequence=Displayed;
CC   Name=Membrane-bound;
CC   IsoId=P04221-1; Sequence=External;
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL: K01357; -; NOT_ANNOTATED_CDS.
DR   PIR: A02164; MHRB.
DR   HSBP: P01842; 7FAB.
DR   InterPro: IPR007110; IG_1like.
DR   InterPro: IPR003597; IG_CL.
DR   InterPro: IPR003006; IG_MHC.
DR   Pfam: PF00047; Ig_4.
DR   SMART: SM00407; IgC1; 2.
DR   PROSITE: PS00835; IG_LIKE; 4.
DR   PROSITE: PS00290; IG_MHC; 3.
KW   Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW   Alternative splicing.
FT   NON_TER          1
FT   DOMAIN          1      106     CH1.
FT   DOMAIN          107     222     CH2.
FT   DOMAIN          223     327     CH3.
FT   DOMAIN          328     458     CH4.
FT   DISULFID          14      14      INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
FT   DISULFID          28      90      BY SIMILARITY.
FT   DISULFID          137     200     INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT   DISULFID          219     219     BY SIMILARITY.
FT   DISULFID          249     308     INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT   DISULFID          296     296     BY SIMILARITY.
FT   DISULFID          356     418     INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT   CARBOHYD          46      46      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD          114     114     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD          212     212     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD          261     261     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD          277     277     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD          284     284     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD          445     445     N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE          458 AA; 49897 MW; F33BD6A3B28E4864 CRC64;

Query Match          10.2%; Score 349; DB 1; Length 458;
Best Local Similarity 24.6%; Pred. No. 3.2e-15;
Matches 116; Conservative 80; Mismatches 167; Indels 108; Gaps 20;

OY   223 FPL-----AFTVEKLTSSGELTMMQAEASSSKSMITFDLKNKSVSVKRVQDPFLQNGKK 277
DB   9 YPLVSCGALTLDNVLVAMGCLARDPLPSVSVFSM-SFK-NNSSTISRTVTRTFPVVKKRGDK 66

```

```

QY 278 -----LPLHLTLPOAL-----POYA 292
DB 67 YWATSGVLPKQVLOGTEBYLVCKQHNSNRDLAVSPVDSLPVSPVFPSPDPS 126
QY 293 GSGNLTALAEAK-TGKLHDEVLNVRATQ-----LQKNLTCGVWGPSPKMLSLKLE 345
DB 127 GSGTRKSRILCQATGSPKQISVSMRLDQOKVESGLTKPVEAFETKGAAPATFSISMLT 186
QY 346 NKEAKYKREKRPVWVLPNPAAGMOCILSDSGVLLSESNIKVLPTWSTPPEPSCKDTHRC 405
DB 187 ITESD-----WL-----SQSLYTCGVDRHG-IFPKNVSM-----SSBC 219
QY 406 PCCPABELLGSPSVFLFPKPKDITLMSRTPBEVTCVVVSHEDPEVKFMVYDGEVFN 465
DB 220 STTSPS-----GLOVFPIASPADT-FLSKARILCLVLTLLTYG-SLNSW-----ASN 268
QY 466 AKT-----KPREQVNSTYRVVSVLTVLHQMVLNGEKYCKVSNKALPAPIEKTISKAKG 520
DB 269 GKALDTHMNTESHPRATFSAMGEASVCAEDMESGQFCTVTHADLPPLKHTTISKSR 328
QY 521 QPRE-POVYTLPPSRBELT-KNOVSLTCLYKGFPSDIAVEMESNGQP--ENNYKTPPV 576
DB 329 VAKHPAVVYVLPAPARQLVRESATVTCVYKGFSPADVFVQWQORQPLSSDKYVTSAPA 388
QY 577 LD--SDGSFFLYSKLTVDKSRMOQGNVFSQVHNEALHNHYTKSLSPG 625
DB 389 PEPQAFGLYTHSTLTVTEEDMNSGETFTCVGHEALPHMVTERTVDKSTG 439

```

RESULT 44

MUC_SUNMU STANDARD; PRT; 457 AA.

```

AC P20768;
DB 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ig mu chain C region.
OS Suncus murinus (House shrew) (Musk shrew).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
NCBI_Taxid=9378;
NCBI_Taxid=9378;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=89232144; PubMed=2497033;
RA Ishiguro H., Ichihara Y., Namikawa T., Nagatsu T., Kurosawa Y.;
RT "Nucleotide sequence of Suncus murinus immunoglobulin mu gene and
RT comparison with mouse and human mu genes.";
RL FEBS Lett. 247:317-322 (1989).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X13920; CAA32113.1; ALT_INIT.
CC PIR; S03961; S03961.
CC HSSP; P01842; 7PAR.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003597; Ig_c1.
CC InterPro; IPR003006; Ig_MHC.
CC Pfam; PF00047; Ig; 4.
CC SMART; SM00407; Igc1; 3.
CC PROSITE; PS00835; IG_Like; 4.
CC PROSITE; PS00290; IG_MHC; 3.
CC Immunoglobulin domain; Immunoglobulin C region; Glycoprotein.
CC NON_TER 1 105 CH1.
CC DOMAIN 106 220 CH2.

```

```

FT DOMAIN 221 326 CH3.
FT DISULFID 327 457 CH4.
FT DISULFID 13 13 INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
FT DISULFID 27 89 BY SIMILARITY.
FT DISULFID 136 200 BY SIMILARITY.
FT DISULFID 217 217 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 248 307 BY SIMILARITY.
FT DISULFID 295 295 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 355 417 BY SIMILARITY.
FT DISULFID 456 456 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 212 212 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 283 283 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 457 AA; 50074 MW; 56C8C086DA4462E9 CRC64;

```

Query Match 10.2%; Score 348; DB 1; Length 457;

Best Local Similarity 23.7%; Pred. No. 3,7e-15; Matches 109; Conservative 90; Mismatches 160; Indels 100; Gaps 18;

```

QY 204 FQKASSIYKKEGEVESEFPPLAFTVEKLTGSGELMWQAE-ASSSKSWITFDLN----- 258
DB 43 FKNSSSI-----SQNIYNPEVFTGKYMATSVLLSTAILQSTDBYITGHTTGE 97
QY 259 KEVSVKRTYQDPKLMGKKLPLHLTL--POLPOYAGSGNLTALAEATGKLHDEVNLV 316
DB 98 KEKV-----ELQVTELPNVAIFVP--PRNSFSGN-----HPTSLI 135
QY 317 MKATOLQKLTCEVA-----GPTPKMLSLKLENKAAYS 352
DB 136 COASGFSRITVMSLQGEPEVQPSLVSVAEAPKSGPTTFVIRLITTE----- 191
QY 353 KKEKPVWVLPNPAAGMOCILSDSGVLLSESNIKVLPTWSTPVEPKSCDHTCPCPAPE 412
DB 192 -----WLSQRE---FTCOALHKG-LTFQKNVSV-----CMGDD 221
QY 413 LLGGSVYVLPFPKPKDITLMSRTPBEVTCVVVSHEDPEVKFMVYDGEVFNATKPRE 472
DB 222 TSTGIVLPLPTFAN-LFTQSAQLTCLVGLATYD-SLISWRONGEALQTHVNISE 279
QY 473 EGVNSTYRVVSVLTVLHQMVLNGEKYCKVSNKALPAPIEKTISKAKQPRE-POVYTL 531
DB 280 SHPNSTFPAKGHASVCREMESGEKFTCTVQHSIDLPKSLSPKPVANDPPEVFLP 339
QY 532 PSRDEL-TKNQVSLTCLYKGFPSDIAVEMESNGQP--ENNYKTPPVLD--SDGSFFLY 586
DB 340 PAQEQCLKRESASTICLVKDFSPDPVFVQWQHGGPVPKHYVTSNPPREPQNPGLYFVH 399
QY 587 SKLTVDKSRMOQGNVFSQVHNEALHNHYTKSLSPG 625
DB 400 SLTVSEKDWSGESGFSQVGHLEALPSTERKAVDKTSG 438

```

RESULT 45

MUC_CANFA STANDARD; PRT; 450 AA.

```

AC P01874;
DB 21-JUL-1986 (Rel. 01, Created)
DB 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig mu chain C region.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
NCBI_Taxid=9615;
NCBI_Taxid=9615;
RN [1]
RP SEQUENCE OF 1-177 (MOO).
RX MEDLINE=80077682; PubMed=117299;
RA McCumber L.J., Capra J.D.;
RT "The complete amino-acid sequence of a canine mu chain.";
RL Mol. Immunol. 16:565-570 (1979).

```

```

RN [12]
RP SEQUENCE OF 178-450 (MOO).
RX MEDLINE=78180587; PubMed=653360;
RA Maaserman R.L., Capra J.D.;
RT "Antino acid sequence of the FC region of a canine immunoglobulin M:
RT interspecies homology for the IGM class.";
RL Science 200;1159-1161(1978).
DR PIR; A93131; MHDG.
DR HSSP; P01857; 1FC1.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00407; IGc1; 2.
DR PROSITE; PS00835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 3.
KM Immunoglobulin domain; Immunoglobulin C region.
FT NON_TER
SQ SEQUENCE 450 AA; 48895 MW; 9D46DDA9D1012F5D CRC64;

Query Match 10.0%; Score 340; DB 1; Length 450;
Best Local Similarity 25.9%; Pred. No. 1,2e-14;
Matches 100; Conservative 76; Mismatches 142; Indels 68; Gaps 15;

QY 278 LPLHLTLPLQAL-----PQYASGN---LTALAEAKTGKHQEVNLVVRATQLOKULT-- 327
DB 102 LPVMTLTPPEVSGFIPRDAFFGGBRKSQLICQASGFSPPQVWSL--RDGQIISGVTTN 159
QY 328 -----CEWGGPTSPFLMLSLKLENKAKVSKREKVVNLNBEAGMOCLLSDSQVLLS 382
DB 160 EVZAZAKSGSPFTTYVTSMULTI-----QEDAWL---SQSYFTCKVHRG-LTFQO 205
QY 383 NIKVLPMTSTPEPKSCDKHTHTCPCPAPELLGSGSVLFPKPKVDTMISRTPEVTCV 442
DB 206 NASSMCTSDQV-----GISITIPPS--PASIFNTKSAKLSTLV 243
QY 443 VDVSHEDEVKFNMYVDGEVHNNAKTKPREQYNSTYRVSVLVYVLMQDMLNGEKYCKV 502
DB 244 TDLATYD-SVTISWIRENGALKTHNTISESHPNCTFSAMGBATVCVEEMESGQFTCTV 302
QY 503 SNKALPAIEKTISSAKG-QPREPOVYTLPPSRDEL-TKNQVSLTCLVKGYSPIADIEN 560
DB 303 THTDLPVALKQITISPKVAAMPSVYVLPSPREQLDRESATLSCLVTGYSPDPVFQW 362
QY 561 ESNQO--PENNYKTPPVLD--SDGSFLYSKLTVDKSRMOQGVNFGSVMEALHNHYT 616
DB 363 VQKGQPVPPDSVYISAPMPPEQAPSLYFAHSILTVSEEMNAGETTYCVVAHESLPKRYT 422
QY 617 QKSLSLSPG-----LQDDETCAE 634
DB 423 ERSVDKSTGKPTLVYVSLVLSPTAGZ 448

RESULT 46
ALC_RABIT STANDARD; PRT; 299 AA.
ID 21-JUN-1986 (Rel. 01, Created)
AC P01879;
DT 21-JUN-1986 (Rel. 01, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DR Ig alpha chain C region (Fragment).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84144059; PubMed=6322114;
RA Knight K.L., Marens C.L., Stoklosa C.M., Schneiderman R.D.;
RT "Genes encoding alpha-heavy chains of rabbit IgA: characterization of
RT cDNA encoding IgA-g, subclases alpha-chains.";
RL Nucleic Acids Res. 12:1657-1670(1984).
CC -1- FUNCTION: Ig alpha is the major immunoglobulin class in body

```

```

CC secretions. It may serve both to defend against local infection
CC and to prevent access of foreign antigens to the general
CC immunologic system.
CC -1- MISCELLANEOUS: This immunoglobulin belongs to the IgA-G subclasses.
CC It was isolated from a rabbit homozygous FOR A2, N60, DE12, 15,
CC F71, G75 heavy chain haplotype.
CC -1- SIMILARITY: Contains 2 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; X00353; CA425100.1; -.
DR PIR; A02174; AHRB.
DR HSSP; P01857; 1FC1.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 2.
DR SMART; SM00407; IGc1; 2.
DR PROSITE; PS00835; IG_LIKE; 2.
DR PROSITE; PS00290; IG_MHC; 2.
KM Immunoglobulin domain; Immunoglobulin C region; Repeat.
FT NON_TER
FT DOMAIN 1 167 IG-LIKE 1.
FT DOMAIN 174 276 IG-LIKE 2.
SQ SEQUENCE 299 AA; 32256 MW; 2512FE3F62B9A223 CRC64;

Query Match 8.9%; Score 305; DB 1; Length 299;
Best Local Similarity 31.6%; Pred. No. 1,2e-12;
Matches 85; Conservative 44; Mismatches 120; Indels 20; Gaps 11;

QY 369 QCLISDSGVLLSNIKVLPTWSTPEPKSCDKHTHTCPCPAP--ELLG--PSVFLP 423
DB 20 QCLGQKSAACHVEYNSV--NESLPVPFPDCCPANSCTCPSSSRRLISGCGPSTLGR 77
QY 424 PKPDTLMISRTPEVTCVVDVSHEDPEVKFNMYVDGEVHNNAKTKPREQYNSTYRV 483
DB 78 PDLGD-LLLRDASLTCTLSGLKNPDAV-FTW--EFTNGNEPVQQAQDLSCYSVSS 133
QY 484 VLTVLHDMNGEKYCKVSNKALPA-PIEKTISSAKGQPREPOVYTLPPSRDELTKN-Q 541
DB 134 VLPSAATWAKARIEFTTVVHPEIDSGSLTATISRGVTF--FQVHLLPPSEELANEQ 191
QY 542 VSLTCLVKGFPYPSDIAVWESNQ--PENNY---KTPPVLDSDGSFLLYSKLTVDKSRW 596
DB 192 VTLTCLVGRFSPKQVLSVMHQEQEVEDSFLWKSMPSSQDKATYATSLRLVPAEDW 251
QY 597 QQGVNFGSVMEALHNHYTKSLSLSPG 625
DB 252 NQGDYSCWVGHGELAEHFTQKTIIDRLAG 280

RESULT 47
MUC_CHICK STANDARD; PRT; 446 AA.
ID 21-JUN-1986 (Rel. 01, Created)
AC P01875;
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DE 15-JUN-1999 (Rel. 38, Last annotation update)
DR Ig mu chain C region.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE OF 80-446 FROM N.A.
RX MEDLINE=83299221; PubMed=6310496;

```

RA Dahan A., Reynaud C.-A., Weill J.-C.;
 RT "Nucleotide sequence of the constant region of a chicken mu heavy
 chain immunoglobulin mRNA."
 RN Nucleic Acids Res. 11:5381-5389(1983).
 RP SEQUENCE OF 1-79 FROM N.A.
 RA Weill J.-C.;
 RL Submitted (MAR-1986) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X01613; CAA25762.1; -
 DR PIR; A02170; MHCH.
 DR HSP; P01857; 1FC1.
 DR InterPro: IPR007110; Ig_1like.
 DR InterPro: IPR003597; Ig_c1.
 DR InterPro: IPR003006; Ig_MHC.
 DR Pfam; PF00047; Ig_3.
 DR SMART; SM00407; Ig1; 1.
 DR PROSITE; PS50835; IG_LIKE; 4.
 DR PROSITE; PS00290; IG_MHC; 3.
 KM Immunoglobulin domain; Immunoglobulin C region.
 FT NON_TER 1 1
 FT DOMAIN 1 105 CH1.
 FT DOMAIN 106 209 CH2.
 FT DOMAIN 210 316 CH3.
 FT DOMAIN 317 427 CH4.
 FT DOMAIN 428 446 C-TERMINAL REGION.
 FT DISULFID 16 16 INTERCHAIN (WITH A LIGHT CHAIN)
 FT DISULFID 27 85 (BY SIMILARITY).
 FT DISULFID 135 190 BY SIMILARITY.
 FT DISULFID 237 296 BY SIMILARITY.
 FT DISULFID 284 284 INTERCHAIN (WITH A HEAVY CHAIN)
 FT DISULFID 344 406 (BY SIMILARITY).
 FT DISULFID 445 445 INTERCHAIN (WITH A HEAVY CHAIN)
 FT CARBOHYD 45 45 (BY SIMILARITY).
 FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 73 73 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 446 AA; 48173 MW; 3CB0CE108949BD17 CRC64;
 Query Match 8.7%; Score 298; DB 1; Length 446;
 Best Local Similarity 27.4%; Pred. No. 5.8e-12;
 Matches 115; Conservative 62; Mismatches 175; Indels 68; Gaps 21;
 Oy 247 SSKSWTTPFLKNEVSVKAVTODPKLQMGKPLHLTLPOALPOYAGSNLTALFAKTG 306
 Db 35 SSIAPTFMDSNNSSVSGMDVI--PKVISG-----PPYAVSR--IQMNSSEG 77
 Oy 307 KLHGEVNLVVRATOLQKLTCEVMP--TSPK--LMLSLKLENK----- 348
 Db 78 KEKQPR--CRAPRGNVEVSVNPGPIPTPNGLPLVTMPPSRKEDEGPFNNASTL 134
 Oy 349 AKVSKREK---VWVLN--PEAGMOCLLSDSGVLLSENIKILPT--NSTPV----- 394
 Db 135 CQTRGRARRPTEVWYNGSPVAATAATTATVGEVAESRIISTSEEMTGTATFSCVNG 194
 Oy 395 EPKSCCKTTC--PPCPABELGGSEVFLPPPKKDTLMISTREVTQVVDVSHEDP-E 451
 Db 195 EWRNTSKRMKCEGLPEVVOODI-----AIRVITSEFVD-IFISKATLTCHVSNMNVADGLE 249

Oy 452 VKFNMYVDGVEVHNAAKTPREQVNSYRVVSVLTVLHQDWLNGKEYCKVSNKALPAPI 511
 Db 250 VSM-WKEGKGKLETLGK-RVLQSGNGLTYVDSVALVCASSNDGGGCKVKNHPLLPEFM 307
 Oy 512 EKTISKAK-GQPREPOVYTLPPSRDELTKNQ-VSLTCLVKGFYPSDIAVEWESNGQ--PE 567
 Db 308 EKKMRKTSANRPPSVYVFPPEPTQLNGNQLSTVTCMAQGFNPPLLFRFMRNGEPLLPQ 367
 Oy 568 NNYKTPPLDLS--DGSFVLVSKLVDSRWQGVFSCSVWHEALHNHYTQKSLSLSPG 625
 Db 368 SOSVTSAPVALENENESVYAVSLVGVABEWAGNVYTCLVGHEALPLQLAQKSDRASG 427
 RESULT 48
 HV02_HETFR
 ID HV02_HETFR STANDARD; PRT; 438 AA.
 AC P23085;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Ig heavy chain C region (Clone 12022) (Fragment).
 OS Heterodontus francisci (Horn shark).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Galeomorphii; Heterodontidae; Heterodontiformes;
 OC Heterodontidae; Heterodontus.
 OX NCBI_TaxID=7792;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=88328985; PubMed=3138109;
 RX Kokubu F., Hinds K., Liltman R., Shambloet M.J., Liltman G.W.;
 RT "Complete structure and organization of immunoglobulin heavy chain
 constant region genes in a phylogenetically primitive vertebrate."
 RL EMBO J. 7:1979-1988(1988).
 CC -1- SIMILARITY: Contains 4 immunoglobulin-like domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X07784; CAA30617.1; -
 DR PIR; S00980; HVKRC2.
 DR HSP; P01842; 7FAB.
 DR InterPro: IPR007110; Ig_1like.
 DR InterPro: IPR003597; Ig_c1.
 DR InterPro: IPR003006; Ig_MHC.
 DR Pfam; PF00047; Ig_4.
 DR SMART; SM00407; Ig1; 2.
 DR PROSITE; PS50835; IG_LIKE; 4.
 DR PROSITE; PS00290; IG_MHC; 3.
 KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
 FT NON_TER 1 1
 FT DOMAIN 1 101 IG-LIKE 1.
 FT DOMAIN 108 202 IG-LIKE 2.
 FT DOMAIN 213 305 IG-LIKE 3.
 FT DOMAIN 315 415 IG-LIKE 4.
 FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 438 AA; 47904 MW; 6E67A9C5BBF7AA7B CRC64;
 Query Match 8.7%; Score 297; DB 1; Length 438;
 Best Local Similarity 27.8%; Pred. No. 6.5e-12;
 Matches 105; Conservative 59; Mismatches 148; Indels 66; Gaps 18;

Qy	291	YASGNITLA-LEAKGKXGHOENLVMBATQ--QNKLTGEWGPISPKMLSLKLENK	347
Db	65	YTRSSQTLTTESEVSGSKTIYCEVR---RGESLMIKELIDCK-GDIVPPTVI-LTGSSS	117
Qy	348	EAKVSKREKEVAVV-----LNPEAGMMOCLSS-----DSGOVL-----LESNIIVLPTWSTPV	394
Db	118	ELITSRRFAVLVCSIIIDFHRESITVSWLKDQGPMSGVTYSPTCEVGNONFSATSRULTVPA	177
Qy	395	EPKSCDKHTCP-----PCPAPELLGSPVFLPPKPKDTLMIISRPE	437
Db	178	GEMFSNTVYTCQVAHQAEVOTSRNITGSQVPCS-----IGDPVYKLLPSEIEQVL-LEATVPT	232
Qy	438	VTCCVVVDVSHEDPEVAFNNYVDCGEVHNAAKTREREOY-----NSTYRVSVLTVLHODW	432
Db	233	LTCV-----VSNAAPYGVNWSV-----TGCKPRKLSEIAVOPGEDSDSVISTVNIQTQW	281
Qy	493	LNCKEYKCKVSNALPAPIEKTISKAGOP-REPOV-VTLPPSRDELTKNOVSLTCLVKG	550
Db	282	LSGAEPFLCVASHODLPTPLRASIHKEEYVDLREPPVSULLPBAEDVSAQORTSLTCLVRG	341
Qy	551	FYPSDIADVEMESNGOEN-NYKTPRPVLDSDG-SFPLYSKLTVDKSRMOOGNFSQVSM	607
Db	342	FSPPREIRIKMTVADKSVNPGNYGNYEVMANENRSPFIYLSLSTLAABEMAGASGYCIVG	401
Qy	608	HEALHNHYTKSLSLSPG	625
Db	402	HEAIPLKTIINRTVNSSG	419

RESULT 49	
HVCM HETFR	
ID_HVCM HETFR	STANDARD;
	PRT; 461 AA

DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig heavy chain C region, membrane-bound form (Clone 3050).
OS *Heterodontus francisci* (Horn shark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Heterodontoidae; Heterodontiformes;
OC Heterodontidae; Heterodontus.
CX NCBI_TaxId=7792;
RN [1]
RN
RP SEQUENCE FROM N.A.
RP TISSUE=Spleen;
RC
RX MEDLINE=86328985; Pubmed=318109;
RA Kokubu F., Hinds K., Litman R., Shambloet M.J., Litman G.W.;
RT "Complete structure and organization of immunoglobulin heavy chain
constant region genes in a phylogenetically primitive vertebrate";
RL EMBO J. 7:1979-1988(1988).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license@isb-sib.ch. (See <http://www.isb-sib.ch/announce> or send an email to license@isb-sib.ch).

DR EMBL: X07781; CAA30614.1; ALT_SEQ.
DR PIR: S01854; HVRKCO.
DR HSBP; P01842; 7PAB.
DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR003597; Ig_c1.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig; 4.
DR SMART: SM00407; IgC1; 2.
DR PROSITE: PSS0835; IG_LIKE; 4.
DR PROSITE: PSS00290; IG_MHC; 3.
DR Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Alternative splicing; Transmembrane.
FT NON TER 1 99 CH1.
FT DOMAIN 1 99

FT	DOMAIN	100	205	CH2.
FT	DOMAIN	206	308	CH3.
FT	DOMAIN	309	418	CH4.
FT	TRANSMEM	438	458	POTENTIAL.
FT	CARBOHYD	164	164	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	200	200	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	245	245	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	275	275	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	374	374	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	411	411	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	415	415	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	437	437	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	461 AA;	50762 MM;	029E91C8BD5DF911 CR664;

Query Match	8.6%;	Score 295;	DB 1;	Length 461;
Best Local Similarly	25.5%;	Pred. No. 9.4e-12;		
Matches 121; Conservative	69;	Mismatches 167;	Indels 118;	Gaps 24

```

OY 208 SSIYKKEGEQEV---SFLAFT-----VEKLTGSELMMQARRASSKEM 251
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 37 TSVMKCKDNEDITTGKLTQYPSVLNKKGTYYSSOLTITESEVSSKLTCEVRGES--VM 94
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 252 ITFPLKNKEVSVKRTQDPKLOMGKCLPLHLTLPOALPOYAGSGMILTALAE----- 303
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 95 I-----KEI-----PDKC---GDKNHPYILITSSSEBITSRFPATVLCGIDHPRES 139
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 304 -----KTGKLHQEVNLVYMRATOLQKNLTCEVMGPTSPKMLSLKLENKEAKVSKREKP 357
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 140 ITVMGLXQGG--HMESGFVT-----SPTGCVNTPFATSRLT-----VPARE-- 179
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 358 VMVLNPEAGMOCCLISDSGOVLBSNIAKVLPTWSPRVEPKSCDKHTHCPCPAPELLGSP 417
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 180 -WPNK---VYTCOVSHOG-VTQSRNT-----TGSQVPCSCND---P 213
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 418 SVFLPFPKPKDLMISRTEVTCVVDVSHEDPEVKFMWYDGVENHNAKTPRREOYNS 477
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 214 VIKLPPRIBQVL-LEATVLTLCV---VSNAPYGVNWSM---TQOKSLKSEIAVQGED 266
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 478 TYRNVSVLTVHODMLNKEYKCKYKSNALAPRIEKTISKAKGP-REPOV-YLTPRPSD 535
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 267 ADVSISTVNISTQMWLSGAEFYCVVNHODLPLPLKASIHKEVKOLREVSAILLSPAD 326
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 536 ELTKNOVSLTLVKGFYPSDIAVENESGOPEN--NYKTPRVLDSL--GSFLEYKCLTYD 592
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 327 VSAQRFLSLTLVAGFFREIFVMTVNDKSVNPNRYNTETVMAANDNSYIITSLSTA 386
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 593 KSRNQOGNVPSCSVNHEALHNHYTKSLSPGLQDETCAEQDGLDGLMTTD 647
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 387 AEEWASGASYSCVGVGHEAI-----PLKINRTVNSKSSPS--DHIIED 426
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 50			
HVCS_HETFR			
ID_HVCS_HETFR	STANDARD:	PRT:	438 AA.
AC	P23087;		
DT	01-NOV-1991 (Rel. 20, Created)		
DT	01-NOV-1991 (Rel. 20, Last sequence update)		
DT	15-JUL-1999 (Rel. 38, Last annotation update)		
DE	Ig heavy chain C region, secreted form (Clone 3050).		
OS	Heterodontus francisci (Horn shark).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;		
OC	Elasmobranchii; Galeomorphi; Heterodontiformes; Heterodontidae; Heterodontus.		
OC	NCL1_TaxID=7792;		
OX	[1]		
RN	SEQUENCE FROM N.A.		
RP	TISSUE=Spleen;		
RC	MEDLINE=88328985; PubMed=3138109;		
RX	Kokubu F., Hinde K., Litman R., Shamloft M.J., Litman G.W.;		
RA	"Complete structure and organization of immunoglobulin heavy chain		
RT	constant region genes in a phylogenetically primitive vertebrate.";		
RL	EMBO J. 7.1979-1988(1988).		
CC	-----		


```

FT DOMAIN 112 207 IG-LIKE 2.
FT DOMAIN 215 317 IG-LIKE 3.
FT DISULFID 26 85 PROBABLE.
FT DISULFID 101 101 INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
FT DISULFID 109 109 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 110 167 PROBABLE.
FT DISULFID 134 191 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 169 169 INTERCHAIN (WITH HEAVY CHAIN IN ANOTHER
FT DISULFID 179 179 SUBUNIT) (POTENTIAL).
FT DISULFID 237 300 PROBABLE.
FT DISULFID 339 339 INTERCHAIN (WITH J CHAIN) (PROBABLE).
FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .).
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .).
FT CARBOHYD 131 131 N-LINKED (GLCNAC. . .).
FT CARBOHYD 205 205 N-LINKED (GLCNAC. . .).
FT CARBOHYD 327 327 P -> S (IN A2M(2) ALLOTYPE).
FT VARIANT 93 93 P -> S (IN A2M(2) ALLOTYPE).
FT VARIANT 102 102 P -> R (IN A2M(2) ALLOTYPE).
FT VARIANT 279 279 F -> Y (IN A2M(2) ALLOTYPE).
FT VARIANT 296 296 D -> E (IN A2M(2) ALLOTYPE).
FT VARIANT 326 326 V -> I (IN A2M(2) ALLOTYPE).
FT VARIANT 335 335 V -> A (IN A2M(2) ALLOTYPE).
SQ SEQUENCE 340 AA; 36508 MW; 9892270756F3276 CRC64;

```

```

Query Match 8.3%; Score 282.5; DB 1; Length 340;
Best Local Similarity 28.3%; Pred. No. 4e-11;
Matches 100; Conservative 47; Mismatches 137; Indels 69; Gaps 17;

Qy 333 PTPSPKLM-LSLKLENKAKVSKR-----EKPYVLINPEAGM-----MOCILSDSG 377
Db 3 PTPSPKPLSLDSTPQDNVVAACLVQGFPOEPLSVWSGQNVARNPPPSQDASGD 62
Qy 378 VLLESNIKVLPWTSPVPEPKS--CD-KHTHCP-----PCPAPBLGGPSVFLPPPK-- 426
Db 63 LYTTSSQLTPLPATQCP-DGKSVTCHVKHATINDSQVTPCPRP-----PPPPCC 110
Qy 427 -----KDTLMISRTPEVTCVVVDVSHEDPEVKENYVDGVEVNAKTKRREQYN 476
Db 111 HPRLSIHRPALEDLLIGSEANLTCTLGL-RDASGATFTWPPSSCK--SAVGPPERRDL 167
Qy 477 STYRVSVLTVLHQQWLNKGEYKCKVSKALPAPIEKTISAKGQPREPOVYTLPPSRDE 536
Db 168 GCYSVSVLPQCAQPMNHGETFTCTAHPBELKTPLTANITKS-GNTPRPVHLLPPPSSE 226
Qy 537 LTKNO-VSLTCLVKGFPYSDIAVENESNGQ--PENNYKTPPVLD-SDG--SFLYSKLT 590
Db 227 LALNELVLTCLAKGFSPKDVLRWLGQSQELPREKVIITMASROBPSSGTTTFATSTLR 286
Qy 591 VDKSRMOQGNVFSQVMEHALNHYTKSLSPS-----LQIDETC 632
Db 287 VAAEDWKKGGDFPSCMVGHGHALPLAFTQKTIDRLAGKPRHVVVSVMAEVDGTC 339

```

```

RESULT 52
HVC3_HETFR STANDARD; PRT; 393 AA.
AC P23086;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DB Ig heavy chain C region (Clone 6121) (Fragment).
OS Heterodontus francisci (Horn shark).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Heterodontioidea; Heterodontiformes;
OC Heterodontidae; Heterodontus.
NCBI_TaxID=7792;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=6832985; PubMed=3138109;
RA Kokubo F., Hinde K., Litman R., Shambloet M.J., Litman G.W.;
RT "Complete structure and organization of immunoglobulin heavy chain
RL EMOB J. 7:1979-1988(1988).
CC -I- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1ib-5ib.ch/announce/
CC or send an email to license@1ib-5ib.ch).
CC
CC -----
CC EMBL; X07782; CAA30615.1; -.
CC PIR; S01852; HYRKC1.
CC HSSP; P01842; 7PAB.
CC InterPro; IPR007110; Ig_Like.
CC InterPro; IPR003597; Ig_c1.
CC InterPro; IPR03006; Ig_MHC.
CC Pfam; PF00047; Ig_3.
CC SMART; SM00407; IgC1; 2.
CC PROSITE; PS00835; IG_LIKE; 3.
CC PROSITE; PS00290; IG_MHC; 3.
KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
FT NON_TER 1
FT DOMAIN 1 157 IG-LIKE 1.
FT DOMAIN 163 260 IG-LIKE 2.
FT DOMAIN 270 370 IG-LIKE 3.
FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 155 155 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 230 230 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 370 370 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 380 380 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 393 AA; 43081 MW; 4E44D076972F18B5 CRC64;

```

```

Query Match 8.2%; Score 281.5; DB 1; Length 393;
Best Local Similarity 26.7%; Pred. No. 5.6e-11;
Matches 116; Conservative 63; Mismatches 164; Indels 91; Gaps 22;

Qy 211 VYKKEGQVPSFPLAFTVEKLTGSGELMWQAEBSASSKSWITFDLKNKEYSVKRTQDP 270
Db 13 VLNKKGYTVO-SSQLTIT-ESEVSSKIYCEVRGES--VWI-----KEI-----PDC 56
Qy 271 KLQWKKGLPLHLTPQLPQYAGSGNLTALAE-----KTGKLHOENVLLV 316
Db 57 K--GDVNHPTVLTITQSSSEITSRPRTVLCSTIIDHPISITVWKLKDGQ-HHESGFTV 112
Qy 317 MRATOLKNTLCEWGFSPSKMLSLKLENKAKVSKREKRPVWVLANPEAGMOCILSDSG 376
Db 113 -----SPTCGNGFSAISRLL-----VPARE--WPTNK---VYTCQVSHQG 149
Qy 377 QVLESNIKVLPWTSPVPEPKSCDKHTTCCPPAPBLGGPSVFLPPPKKDTLMISRTTP 436
Db 150 -VTQSRNI-----TGSQVPCSGND-----PVIKLPLPSIEQVL-LEATV 186
Qy 437 EVTCVVVDVSHEDPEVKENYVDGVEVNAKTKRREQYNSTYRVSVLTVLHQQWLNK 496
Db 187 TLNCT--VSNAPGVNVSW--TQEQSKLSSEIAVPGEDADSVISVTWISTQAMLSGA 240
Qy 497 EYCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPS 554
Db 241 EFYCVNVHQQDLPPLTPRSIHKEVYKDKRERSVSLTSPADVSQRSLTCLVKRGSFR 300
Qy 555 DIAVESNGQOPEN-NYKTTTPPVLDSD-GSFLYSKLTVDKSRMOQGNVFSQVMEHAL 611

```

Db 301 ELPVKTINDKSVNPGNYKNTVEAMENDNSYFYISLSIAAEWASGASVCVGHENI 360
QY 612 HHNYTOKSLSPG 625
Db 361 PLKINRTYVKNSSG 374

RESULT 53
ALCI HUMAN STANDARD; PRT; 353 AA.
ID 1
AC P01876; 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
GN IGH1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84130179; PubMed=6421489;
RA Flanagan J.G., Lettrant M.-P., Rabbits T.H.;
RT "Mechanisms of divergence and convergence of the human immunoglobulin
RT alpha 1 and alpha 2 constant region gene sequences."; Cell 36:681-688(1984).
RL [2]
RN SEQUENCE (MYELOMA PROTEIN BUR), AND DISULFIDE BONDS.
RX MEDLINE=79151016; PubMed=107164;
RA Putnam F.W., Liu Y.-S.V., Low T.L.K.;
RT "Primary structure of a human IgA1 immunoglobulin. IV. Streptococcal
RT IgA1 protease, digestion, Fab and Fc fragments", and the complete
RT amino acid sequence of the alpha 1 heavy chain."; J. Biol. Chem. 254:2865-2874(1979).
RL [3]
RN SEQUENCE (MYELOMA PROTEIN TRO).
RX MEDLINE=76023781; PubMed=809331;
RA Kratzin H., Altevogt P., Ruban E., Kortt A., Staroscek K.,
RA Hilschmann N.;
RT "The primary structure of a monoclonal IgA-immunoglobulin (IgA Tro.),
RT II. The amino acid sequence of the H-chain, alpha-type, subgroup III;
RT structure of the complete IGA-molecule."; Hoppe-Seyler's Z. Physiol. Chem. 356:1337-1342(1975).
RL [4]
RN DISULFIDE BONDS.
RX MEDLINE=80114124; PubMed=393607;
RA Yang C.-Y., Kratzin H., Gotz H., Hilschmann N.;
RT "Rule of antibody structure. Primary structure of a human monoclonal
RT IgA-immunoglobulin (myeloma protein Tro). VII. Purification and
RT characterization of the disulfide bridges."; Hoppe-Seyler's Z. Physiol. Chem. 360:1919-1940(1979).
RL [5]
RN REVIEW.
RX MEDLINE=91054387; PubMed=2241915;
RA Kerr M.A.;
RT "The structure and function of human IgA."; Biochem. J. 271:285-296(1990).
RL [1]- FUNCTION: Ig alpha is the major immunoglobulin class in body
CC secretions. It may serve both to defend against local infection
CC and to prevent access of foreign antigens to the general
CC immunologic system.
CC [1]- SUBUNIT: Monomeric or polymeric.
CC [1]- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: J00220; AAC82528.1; ALT_INIT.
DR PIR: A22360; A1HU.
DR PDB: 1IGA; 15-JUN-99.
DR Genew: HGNC:5478; IGH1.
DR MIM: 146900; -.
DR GO: GO:0003823; F:antigen binding; NAS.
DR GO: GO:0006955; P:immune response; NAS.
DR InterPro: IPR007110; IG-like.
DR InterPro: IPR003597; IG cl.
DR InterPro: IPR003006; IG_MHC.
DR Pfam: PF00047; Ig; 3.
DR SMART: SM00407; IGH1; 2.
DR PROSITE: PS50835; IG LIKE; 3.
DR PROSITE: PS00290; IG MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Polymorphism; 3D-structure; Repeat.
FT NON_TER 1 1
FT DOMAIN 6 98
FT FT 125 220
FT DOMAIN 228 330
FT FT 14 14
FT DISULFID 26 85
FT FT 77 101
FT DISULFID 122 122
FT DISULFID 123 180
FT DISULFID 147 204
FT FT 182 182
FT DISULFID 192 192
FT FT 250 313
FT DISULFID 352 352
FT CARBOHYD 105 105
FT CARBOHYD 111 111
FT CARBOHYD 113 113
FT CARBOHYD 119 119
FT CARBOHYD 121 121
FT CARBOHYD 144 144
FT CARBOHYD 340 340
FT VARIANT 176 176
FT FT 163 165
FT CONFLICT 176 176
FT CONFLICT 190 190
FT CONFLICT 227 227
FT CONFLICT 231 231
FT CONFLICT 290 290
SQ SEQUENCE 353 AA; 37654 MW; EBA11ECB7E85DB21 CRC64;

Query Match 8.2%; Score 280.5; DB 1; Length 353;
Best Local Similarity 26.8%; Pred. No. 5.6e-11;
Matches 93; Conservative 49; Mismatches 132; Indels 73; Gaps 13;

QY 305 TGKLGQEVNLVVMRATQ--LQKNLTCEVNGPSTPKMLSLKLENKEAKVSRKPVWVLN 362
Db 60 SGLDYTTSSQULPAPTOCLAGKSVTCVYKHTNPSQDVTVP----- 100
QY 363 PEAGMGOCLLSDSGOVLLESNIKVLPTWSTPEVPSKCDKHTCPCPAPBELLGGPSVLF 422
Db 101 -----CPVPSTPTPT-----SPSTPTPTPSGCC-HPRLSLH 130
QY 423 PPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNNYVDSGVENAKTKRREGQYNSTRYV 482
Db 131 RPALEED-LLTSEANLTCTLTGL-RDASGVFTFTWPSGK--SAVQGPREDLGGCVSVS 186
QY 483 SVLTATVHODMNGEKYCKVKSNKALPAPIETGISAKGQPREPOVYTLPPSRDELTRNQ- 541
Db 187 SVLPCCAPPMWNGKFTTAAVPSKSTPLTTLTKS-GNTRPPEVHLPPSEELALNEL 245
QY 542 VSLTGLVGFYPSDIAVWESNGQ--PENNYKTPPVLD-SDG--SFFLYSKLTIVDKSRW 596
Db 246 VTLTGLAAGFSPKDVLVWMLGSGDELPREKVLTVNASRQEPGQTTTFAVTSILKVAEDW 305

```

QY 597 OQGNVSCSVHHEALHNHYTKSLSPG-----LQDSTC 632
DB 306 KKGDTFSCVGHLEALPLAFTQKTIDRLAGKPTHVNVSVMAEVDGTC 352

RESULT 54
ALCL_GORGO STANDARD; PRT; 353 AA.
ID ALCL_GORGO
AC P20758;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig alpha-1 chain C region.
GN IGHAI.
OS Gorilla gorilla gorilla (Lowland gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Catarrhini; Homiidae; Gorilla.
OX NCBI_TaxId=9595;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph node; PubMed=2506527;
RX MEDLINE=89386006;
RA Kawamura S., Omoto K., Ueda S.;
RT "Nucleotide sequence of the gorilla immunoglobulin alpha 1 gene.";
RL Nucleic Acids Res. 17:6732-6732(1989).
CC -1- FUNCTION: Ig alpha is the major immunoglobulin class in body
CC and to prevent access of foreign antigens to the general
CC immunologic system.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X15045; CAA33147.1; ALT_INIT.
CC HSSP: P01810; 2FBJ.
DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR003597; Ig_c1.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig_3.
DR SMART: SM00407; Ig_c1; 2.
DR PROSITE: PS50835; IG_LIKE; 3.
DR PROSITE: PS00290; IG_MHC; 1.
KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
FT NON_TER 1 1
FT DOMAIN 6 98
FT DOMAIN 125 220
FT DOMAIN 228 330
FT DISULFID 14 14
FT DISULFID 26 85
FT DISULFID 77 101
FT DISULFID 122 122
FT DISULFID 123 180
FT DISULFID 147 204
FT DISULFID 182 182
FT DISULFID 192 192
FT DISULFID 250 313
FT DISULFID 352 352
FT CARBOHYD 144 144
FT CARBOHYD 340 340
SQ SEQUENCE 353 AA; 37755 MW; 4820B8DB02AC7514 CMC64;

Query Match
Best Local Similarity 28.1%; Score 278; DB 1; Length 353;
Pred. No. 8.1e-11;

```

```

Matches 101; Conservative 47; Mismatches 143; Indels 68; Gaps 17;
QY 333 PTSEKLM-LSLKENKAKVSKR-----EKPVWLNDEAGW-----WQCLLSDSQ 377
DB 3 PTSPKVFPLSLCSTQDPDGVVACLVGFPQPEPLSTVWSESGGVYARNPPESQDASD 62
QY 378 VLBSNKKVLPSTVPEPKS--CDKTH-----TCPPCAPELLGGSVF 420
DB 63 LYTTSSQTLPLATQCP-DGKSVTCHVNHYNTPSQDVTVPGRVSTPTPTSPSPPTPS-- 119
QY 421 LFPF-----KPK-DTLMISRTPEVTCVVVDVSHEDPEVKRWYVDGVEVNATKTP 470
DB 120 --PPCCHPLSLHRLALDILLGSEANLTCTLTGL-RDASGVFTWTPSSGK--SAVECP 174
QY 471 REEQNSTYRVVSVLTVADHWLNGEKVKVSKALPAPIEKTIKSKAQPREQVYTL 530
DB 175 PERLQCGYSVSLTPGCAPPMHNGKFTCTAAVPEKTLTLTLSS-GNMPPEVHL 233
QY 531 PPSRDELTKQ-VSLTCLVKGFPSPDIIVEMESNQ--PENNYKTPPVLD-SDG--SFF 584
DB 234 PPSSEELALNELVTLTCLARFSPKDVLRWLGSSQELPREKYLTMASRQEPSSQTTFA 293
QY 585 LYSKLTVDKSRMOQGNVSCSVHHEALHNHYTKSLSPG-----LQDSTC 632
DB 294 VTSILRVAAEDWKKGDTFSCVGHLEALPLAFTQKTIDRLAGKPTHVNVSVMAEVDGTC 352

RESULT 55
HVC1_HETFR
ID HVC1_HETFR STANDARD; PRT; 370 AA.
AC P23084;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain C region (Clone 6125) (Fragment).
OS Heterodontus francisci (Horn shark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphi; Heterodontidae; Heterodontiformes;
OC Heterodontidae; Heterodontus.
OX NCBI_TaxId=7792;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=88328985; PubMed=3138109;
RX Kokubu F., Hinds K., Litman R., Shambloct M.J., Litman G.W.;
RT "Complete structure and organization of immunoglobulin heavy chain
RT constant region genes in a phylogenetically primitive vertebrate.";
RL EMO J. 7:1979-1988(1988).
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X07783; CAA30616.1; -.
CC PIR: S01851; HVKACS.
DR HSSP: P01857; 1FCL.
DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR003597; Ig_c1.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig_3.
DR SMART: SM00407; Ig_c1; 2.
DR PROSITE: PS50835; IG_LIKE; 3.
DR PROSITE: PS00290; IG_MHC; 1.
KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
FT NON_TER 1 1
FT DOMAIN 40 134
FT DOMAIN 145 237
FT DOMAIN 247 347
FT CARBOHYD 98 98
SQ SEQUENCE 370 AA; 37755 MW; 4820B8DB02AC7514 CMC64;
N-LINKED (GLCNAC. . .) (POTENTIAL).

```

	FT	CARBOHYD	132	132	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	132	177	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT	CARBOHYD	343	343	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT	CARBOHYD	347	347	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT	CARBOHYD	357	357	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
SEQUENCE		370 AA;	40586 MW;	32833A6EF7AEDB0	CRC64;	
Query Match			7.8%;	Score 267;	DB 1; Length 370;	
Best Local Similarity			32.4%;	Pred. No. 4, 4e-10;		
Matches			71;	Conservative 40;	Mismatches 86; Indels 22; Gaps 8	
Qy	417	PSVFLPPKPKDITLMISRTPEVTCVVDVSHDEPEVKFMYDGVGVHNAKTKRPEEQY-	475			
Db	145	PAIKLLPSPSIEQVL-LEATVTLVTCV---VSNAPYGVNWSM-----TQEQKPLKSEIA	192			
Qy	476	-----NSTRRVVSVLTVLHODMLNGEKYKCKVSNNKLLPAPIETKISK-AKGQPREPOVYTL	530			
Db	193	VQPGEDPDSVISTVDISTQAWLSEAVFYCVSVHQDLPTLRDSIHKEAWKDLPRESVSVL	252			
Qy	531	PPSRBELTKNQ-VSLTCLVKGFPSPDIAVEMWSNQPER-NYKTTTPVLDD-DGSFLY	586			
Db	253	LPPEAEISAERELSTLCTLVGRGSPREIFVKMTVNNKSVNPGNYKOTEVAENDKSSFFLY	312			
Qy	587	SKLTVDKSRMOQGNVFCSSVMHEALHNHTQKSLSPG	625			
Db	313	SLLSIAAEEMWASGASVSCVGHGHALPLKIINTVKNSSG	351			

```

RESULT 56
MUCM ICTPU
ID _MUCM ICTPU STANDARD: PRT: 481 AA.
AC P23735;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-NOV-2004 (Rel. 43, Last annotation update)
DE Ig mu chain C region membrane-bound form.
OS Ictalurus punctatus (Channel catfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Siluriformes;
OC Ictaluridae; Ictalurus.
OX NCBI_TaxID=7998;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90384824; Pubmed=2119496;
RA Wilson M.R., Marcuz A., van Ginkel F., Miller N.W., Clem L.W.,
RA Middleton D., Watt G.W.;
RT "The immunoglobulin M heavy chain constant region gene of the channel
RT catfish, Ictalurus punctatus: an unusual mRNA splice pattern produces
RT the membrane form of the molecule."
RL Nucleic Acids Res. 18:5227-5233(1990).
CC -I- MISCELLANEOUS: During differentiation, B lymphocytes switch from
CC expression of membrane-bound IgM to secretion of IgM. The mu
CC chains of membrane and secreted IgM differ in their C-terminal
CC segments.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; X52617; -; NOT_ANNOTATED_CDS.
DR HSSP; P01857; 1FCL.
DR InterPro; IPR007110; Ig-1Ike.
DR InterPro; IPR003597; Ig_C1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00407; IgC1; 1.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 2.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;

```

[illegible]

Query Match	Similarity	7.8*	Score 265.5	DB 1	Length 481
Best Local	Similarity	25.7*	Pred. No. 7.7e-10		
Matches	93	Conservative	59	Mismatches	149
				Indels	61
				Gaps	17
QY	294	SGNTLTALAEAKTG-----	KLHGEVNLVVMRATOLQKNLTCEYW--	GPTSPK	337
	:	:	:	:	:
	:	:	:	:	:
DB	78	NGNKKFCFCEVNGNGSGSDASIQKREVERELHSLTLTPQTEINDGRTFPCLATPSPK			137
QY	338	LMLSLKLENKEAKVSKREKPVWVLPNPEAGWQCLISDSGQVLLSNIKVLPFTWSTPEPK			397
	:	:	:	:	:
	:	:	:	:	:
DB	138	-SHTFPKLTLEKTDISNKKVEN-IYSONKGNFTAI-----	SVLEISAS---	EMTSSTSPV	186
QY	398	SCD---KTHTC--PPCAPELLGSPSVLPFPKKKDTLMISRTBEYVCVVVDVSHDEBEV			452
	:	:	:	:	:
	:	:	:	:	:
DB	187	KCEPQOKNHNVPKEASVAPGDTKOPQVKTIGSPSTD-LILIRAGOLEC---	RAEGDTGF		241
QY	453	K-FMYVDGVEVHAHNAKTPREEQVNSTYRVVSVLT-VLHODMLNGKEKCKVSKNALPAP			510
	:	:	:	:	:
	:	:	:	:	:
DB	242	KSIMMLIGNRISLSLN-----	SKTIVSLQTHIGFEEMINGTEPICVEHHAFTQQ		294
QY	511	IEK-TISKANGQPREPOVYTLPPSRDELITKNQVSLTCLVKGFPYSDIAVEMESN----			563
	:	:	:	:	:
	:	:	:	:	:
DB	295	YEKVTFKRENGNPFEPKVIYLLAP--	PESSGSEVTLTCVXKDFYPREAVSVMLVNDQYEE		352
QY	564	--GQPENNYKTPPYVLDSDGSFFLYSKLTVDKSKMOQGNVSSCSVMHETALNNHTQKLS			621
	:	:	:	:	:
	:	:	:	:	:
DB	353	VVGYEQN---TTAVIDRNMLFSYSQLLITKTDAPMNSGVSFCLVYHESIKDCVARIHSR			408
QY	622	LS	623		
	:	:	:	:	:
	:	:	:	:	:
DB	409	IA	410		
RESULT 57					
ALC MOUSE					
ID _ALC MOUSE	STANDARD;	PRT;	344	AA.	
AC	P01878;				
DT	21-JUL-1986	(Rel. 01, Created)			
DT	21-JUL-1986	(Rel. 01, Last sequence update)			
DT	15-MAR-2004	(Rel. 43, Last annotation update)			
DE	IG alpha chain C region.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A. (MYELOMAS ABB48 AND J558).				
RX	MEDLINE=81261947; PubMed=6790349;				
RA	Auffray C., Nageotte R., Sikorav J.-L., Heidmann O., Rougeon F.;				
RT	"Mouse Immunoglobulin A: nucleotide sequence of the structural gene				
	for the alpha heavy chain derived from cloned cDNAs."				

```

RL Gene 13:365-374(1981).
RN [2]
RP SEQUENCE OF 1-213 (MOPC 47A).
RX MEDLINE=80049769; PubMed=115869;
RA Robinson E.A.; Appella E.;
RT "Amino acid sequence of a mouse myeloma immunoglobulin heavy chain
  (MOPC 47 A) with a 100-residue deletion.";
RL J. Biol. Chem. 254:11418-11430(1979).
RN [3]
RP SEQUENCE OF 1-254 AND 291-344 (M511).
RX MEDLINE=81054880; PubMed=6776528;
RA Robinson E.A.; Appella E.;
RT "Complete amino acid sequence of a mouse immunoglobulin alpha chain
  (MOPC 511).";
RL Proc. Natl. Acad. Sci. U.S.A. 77:4909-4913(1980).
CC -I- FUNCTION: Ig alpha is the major immunoglobulin class in body
  secretions. It may serve both to defend against local infection
  and to prevent access of foreign antigens to the general
  immunologic system.
CC -I- MISCELLANEOUS: The final C-region domain is deleted from Ref.2
  chain. It was isolated from a myeloma protein that contains 1
  light and 1 heavy chain per molecule, linked by a disulfide bond.
  In contrast, normal mouse IgA molecules contain 2 light and 2
  heavy chains and lack a light-heavy chain disulfide bond.
CC -I- MISCELLANEOUS: M511 chain was isolated from a myeloma protein that
  binds phosphorylcholine.
CC -I- MISCELLANEOUS: M511 sequence was compared with that of mouse MOPC
  47A, and a genetic mechanism for the deletion of the CH3 domain of
  the mutant chain is proposed.
CC -I- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL, D11468; BA002026.1; ALT_INTR.
DR PIR, A91479; AHMS.
DR HSSP, P01810; 2FBJ.
DR InterPro; IPR007110; IG-1Lk.
DR InterPro; IPR003597; IG_C1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig_3.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 2.
DR Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
KT NON TER
FT DOMAIN 1 6 99
FT 116 206 IG-LIKE 1.
FT DOMAIN 219 321 IG-LIKE 3.
FT DISULFID 26 84 BY SIMILARITY.
FT DISULFID 76 100 BY SIMILARITY.
FT DISULFID 114 171 BY SIMILARITY.
FT DISULFID 138 195 BY SIMILARITY.
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .).
FT CARBOHYD 101 101 O-LINKED.
FT CARBOHYD 331 331 O-LINKED.
FT VARIANT 2 2 S -> A (IN MOPC 47A).
FT VARIANT 18 18 S -> C (IN MOPC 47A AND M511).
FT VARIANT 67 67 N -> S (IN MOPC 47A AND M511).
FT VARIANT 73 73 A -> T (IN MOPC 47A).
FT VARIANT 112 112 P -> G (IN M511).
FT VARIANT 135 135 S -> Q (IN MOPC 47A AND M511).
FT VARIANT 141 141 N -> D (IN MOPC 47A AND M511).
FT VARIANT 168 168 O -> E (IN MOPC 47A).
FT VARIANT 212 213 VT -> SQ (IN MOPC 47A).
FT VARIANT 235 235 E -> G (IN M511).
FT VARIANT 255 290 MISSING (IN M511).
FT VARIANT 295 T -> D (IN M511).

```

```

FT VARIANT 301 301 Q -> G (IN M511).
FT VARIANT 329 329 N -> Q (IN M511).
FT VARIANT 331 331 S -> N (IN M511).
SQ SEQUENCE 344 AA; 36876 MW; 3694CF99B19A988 CRC64;
Query Match 7.3%; Score 248.5; DB 1; Length 344;
Best Local Similarity 25.5%; Pred. No. 6.1e-09;
Matches 93; Conservative 56; Mismatches 145; Indels 71; Gaps 14;
QY 282 LTLPQALP-----QYASGNLTLLAEKTKGLHDEVNLYVMRAIQOLNLTCEV 330
  |||||
  11 LTPPALSSDPVIGICLIHDYFPSPGTMNVTW-GKSGDITTVNFPPLASGGRYTMNQL 69
QY 331 WGFPSPLMLSLKLENKAVKSKREKPVWLNPNAGWMOCLSDSGVLLSNIKTVL-PT 389
  |||||
  70 TLPR-----VECBGEVSKSVQVHDSNPVDELVNCGPT 104
QY 390 WSTPVEPKSCDKHTPCPCAPALLGSPVFLPPPKKDTLMSRPEVTGVVDVSHED 449
  |||||
  105 PPPPI-----TIPSCQ-----PSLSQRPALPD-LTIGSDASITCTLNGLRNPE 147
QY 450 PEYKFNMYVDGVVHNNAKTPREQVNST---YRVSVLVTLVQDMVNGEKYCKNSKA 506
  |||||
  148 GAV-FTV-----EPSTGKDAVQKKAQVNSGCVSVSLGCAERMSGASFCTVTHPE 201
QY 507 LPAPIEKTIKAKQGPPEPVVTLPPSRDELTKNQ-VSLTCLVKGFPSPDIAYEWESNGQ 565
  |||||
  202 -SGLTGTITAKVTVNTPRPVHLLPPESEIALNELSLTCLVRAFPKRVLRMLHGNE 260
QY 566 ---PENNYKTPPYLSDG--SFFLYSKLTVDKSRMOQGVNFGSVNHEALHNHYTKSL 620
  |||||
  261 ELSPESVLYPELPKEPEGATVTVLVSATETWKGQVSGVMGHEALPMNFQXTI 320
QY 621 SLSPG 625
DB 321 DRUG 325

```

RESULT 58

SHS1_MOUSE STANDARD; PRT; 513 AA.

ID SHS1_MOUSE AC P97797; O08907; O35924; O88555; O88556; P9796; Q08559; Q09X57;

AC Q9WTN4; DT 10-OCT-2003 (Rel. 42, Created) DT 10-OCT-2003 (Rel. 42, Last sequence update) DT 10-OCT-2003 (Rel. 42, Last annotation update) DE Protein-tyrosine phosphatase non-receptor type subfamily 1 precursor DE (SHP substrate-1) (SHPs-1) (inhibitory receptor SHP-1) (Signal-regulatory protein alpha-1) (SIRP-alpha-1) (msIRP-alpha1) (MyD-1 antigen) (Brain Ig-1-like molecule with tyrosine-based activation DE motif) (Bit) (p84). DE CN PRPNS1 OR SHPS1 OR SIRP OR MYD1 OR BIT. OS Mus musculus (Mouse). OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. OX NCBI_Taxid=10090; [1] RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2). RC TISSUE=Brain; RX MEDLINE=97223399; PubMed=9070220; RA Yamao T.; Matozaki T.; Amano K.; Matsuda Y.; Takahashi N.; Ochi F.; RA Fujioke Y.; Kasuga M.; RT "Mouse and human SHP-1: molecular cloning of cDNAs and chromosomal RT localization of genes."; Biochem. Biophys. Res. Commun. 231:61-67(1997). RN [2] RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANTS ALA-29; ARG-67; ARG-91; RP THR-96; SER-128; PRO-194 AND ASN-224. RC STRAIN=BA1B/C; TISSUE=Brain; RC MEDLINE=97230468; PubMed=9073522; RA Ohnishi H.; Kubota M.; Sano S.-I.; RT "BIT (Bit) maps to mouse chromosome 2."; Genomics 40:504-506(1997).

```

RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), SEQUENCE OF 32-53 AND 422-433
RP (ISOFORM 2), VARIANTS ALA-29; ARG-91; THR-96 AND SER-128,
RP N-GLYCOSYLATION, AND TISSUE SPECIFICITY.
RA STRAIN=BALB/c; TISSUE=Brain, and Cerebellum;
RX MEDLINE=98012243; PubMed=9348339;
RA Comu S., Wang W., Olinick S., Ishwad P., Mi Z., Hempel J., Watkins S.,
RA Lagenaur C.F., Narayanan V.;
RT "The murine P84 neural adhesion molecule is SHPS-1, a member of the
RT phosphatase-binding protein family.";
RL J. Neurosci. 17:8702-8710(1997).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), VARIANTS ALA-29; ARG-67;
RP PRO-74; ALA-83; 86-TYR-VAL-87; ILE-90; ARG-91; THR-96; ALA-114;
RP ILE-118; SER-128; PRO-194; ASN-224; PHE-351 AND ASP-365,
RP N-GLYCOSYLATION, PHOSPHORYLATION ON TYROSINE RESIDUES, INTERACTION
RP WITH PTN6, AND TISSUE SPECIFICITY.
RC STRAIN=C57BL/6; TISSUE=Fetal thymus;
RX MEDLINE=98380500; PubMed=9712903;
RA Veilleux A., Thibaudan E., Latour S.;
RT "High expression of inhibitory receptor SHPS-1 and its association
RT with protein tyrosine phosphatase SHP-1 in macrophages.";
RL J. Biol. Chem. 273:22719-22728(1998).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3), VARIANTS ALA-29; ARG-67;
RP PRO-74; ALA-83; 86-TYR-VAL-87; ILE-90; ARG-91; THR-96; ALA-114;
RP ILE-118; SER-128; PRO-194; ASN-224; PHE-351 AND ASP-365, AND
RP N-GLYCOSYLATION.
RC STRAIN=129/SvJ; TISSUE=Brain, and Liver;
RX MEDLINE=20053880; PubMed=1058583;
RA Sano S.-I., Ohnishi H., Kubota M.;
RT "Gene structure of mouse Birt/SHS-1.";
RL Biochem. J. 344:667-675(1999).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANTS ALA-29; ARG-67;
RP PRO-74; ALA-83; 86-TYR-VAL-87; ILE-90; ARG-91; THR-96; ALA-114;
RP ILE-118; SER-128; PRO-194; ASN-224; PHE-351 AND ASP-365.
RC STRAIN=ILIS, and ISS;
RX MEDLINE=21363810; PubMed=11471062;
RA Ehringer M.A., Thompson J., Conroy O., Xu Y., Yang F., Canniff J.,
RA Beeson M., Gordon L., Bennett B., Johnson T.E., Sikela J.M.;
RT "High-throughput sequence identification of gene coding variants
RT within alcohol-related OTRs.";
RL Mamm. Genome 12:657-663(2001).
RN [7]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT PRO-194.
RC TISSUE=Placenta;
RA Wang H., Chen Z., Ullrich A.;
RT "Epidermal growth factor-induced association of SHP2 with mouse SHP-
RT alpha1.";
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [8]
RP FUNCTION, TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RX MEDLINE=90152134; PubMed=2303162;
RA Chuang W., Lagenaur C.F.;
RT "Central nervous system antigen P84 can serve as a substrate for
RT neurite outgrowth.";
RL Dev. Biol. 137:219-232(1990).
RN [9]
RP N-GLYCOSYLATION, PHOSPHORYLATION BY JAK2 IN RESPONSE TO GROWTH
RP HORMONE, AND INTERACTIONS WITH JAK2 AND PTPN11.
RX MEDLINE=98175985; PubMed=9507023;
RA Stoege M.R., Wang H., Ullrich A., Carter-Su C.;
RT "Growth hormone regulation of SHP and SHP-2 tyrosyl phosphorylation
RT and association.";
RL J. Biol. Chem. 273:7112-7117(1998).
RN [10]
RP INTERACTION WITH CD47, AND TISSUE SPECIFICITY.
RX MEDLINE=99091586; PubMed=9872987;
RA Jiang P., Lagenaur C.F., Narayanan V.;
RT "Integrin-associated protein is a ligand for the P84 neural adhesion
RT molecule.";
RL J. Biol. Chem. 274:559-562(1999).

```

```

RN [11]
RP INTERACTION WITH FGR.
RX MEDLINE=20130295; PubMed=10662797;
RA Gresham H.D., Dale B.M., Potter J.W., Chang P.W., Vines C.M.,
RA Lowell C.A., Lagenaur C.F., Willman C.L.;
RT "Negative regulation of phagocytosis in murine macrophages by the Src
RT kinase family member, Fgr.";
RL J. Exp. Med. 191:515-528(2000).
RN [12]
RP FUNCTION: Immunoglobulin-like cell surface receptor for CD47. Acts
RP as docking protein and induces translocation of PTPN6,
RP PTPN11 and other binding partners from the cytosol to the
RP plasma membrane. Supports adhesion of cerebellar neurons, neurite
RP outgrowth and glial cell attachment. May play a key role in
RP intracellular signaling during synaptogenesis and in synaptic
RP function. Involved in the negative regulation of receptor tyrosine
RP kinase-coupled cellular responses induced by cell adhesion, growth
RP factors or insulin. Mediates negative regulation of phagocytosis,
RP mast cell activation and dendritic cell activation. CD47 binding
RP prevents maturation of immature dendritic cells and inhibits
RP cytokine production by mature dendritic cells (By similarity).
RN [13]
RP SUBUNIT: Binds PTPN11 when tyrosine-phosphorylated, except in
RP macrophages, where it primarily binds PTPN6. Binds GRB2 in vitro.
RP Binds FGR. Binds JAK2 irrespective of its phosphorylation status
RP and forms a stable complex. Binds SCAP1 and/or SCAP2. The
RP resulting complex recruits Fyb. Binds PIK2B (By similarity).
RN [14]
RP ALTERNATIVE PRODUCTS: Type I membrane protein.
RP Event=Alternative splicing; Named isoforms=3;
RP Comment=Additional isoforms seem to exist;
RP Name=1; Synonyms=a;
RP IsoId=p97797-1; Sequence=Displayed;
RP Name=2; Synonyms=a', Large; VSP_007032;
RP IsoId=p97797-2; Sequence=VSP_007031;
RP Name=3; Synonyms=b, Small;
RP IsoId=p97797-3; Sequence=VSP_007031;
RN [15]
RP TISSUE SPECIFICITY: Highly expressed in cerebral cortex, brain,
RP spinal cord, cerebellum and spleen, and at much lower levels in
RP kidney, thymus, heart, lung and liver. Within the cerebellum,
RP glomeruli in the granule cell layer. Detected in neurons of the
RP hippocampus and dentate gyrus, and in olfactory bulb. Not detected
RP in Purkinje cells. Highly expressed in the plexiform layers, optic
RP fiber layer and the outer segments of the photoreceptor layer in
RP the retina. Highly expressed in macrophages. Isoform 3 is detected
RP at very low levels in all tissues tested.
RN [16]
RP DEVELOPMENTAL STAGE: Highly expressed in the CNS of embryos from
RP day 7 to 17.
RN [17]
RP PTM: N-glycosylated.
RN [18]
RP PTM: Phosphorylated on tyrosine residues.
RN [19]
RP SIMILARITY: Contains 2 immunoglobulin-like C1-type domains.
RN [20]
RP SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
RN [21]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RP between the Swiss Institute of Bioinformatics and the EMBL outstation -
RP the European Bioinformatics Institute. There are no restrictions on its
RP use by non-profit institutions as long as its content is in no way
RP modified and this statement is not removed. Usage by and for commercial
RP entities requires a license agreement (See http://www.isb-sib.ch/announce/
RP or send an email to license@isb-sib.ch).
RN [22]
RP EMBL, D87967; BAA13520.1; -
RP EMBL, D87968; BAA13521.1; -
RP EMBL, D85785; BAA20376.1; -
RP EMBL, D85785; BAA20376.1; -
RP EMBL, AF072543; AAC24886.1; -
RP EMBL, AF072544; AAC24887.1; -
RP EMBL, AB024507; BAA89290.1; -
RP EMBL, AB024508; BAA89290.1; JOINED.
RP EMBL, AB024509; BAA89290.1; JOINED.
RP EMBL, AB024510; BAA89290.1; JOINED.
RP EMBL, AB024502; BAA89290.1; JOINED.
RP EMBL, AB024503; BAA89290.1; JOINED.
RP EMBL, AB024504; BAA89290.1; JOINED.
RP EMBL, AB024505; BAA89290.1; JOINED.

```

DR EMBL; AB024506; BAA89290.1; JOINED.
 DR EMBL; AB018194; BAA76555.1; -.
 DR EMBL; AB024507; BAA89289.1; -.
 DR EMBL; AB024500; BAA89289.1; JOINED.
 DR EMBL; AB024501; BAA89289.1; JOINED.
 DR EMBL; AB024502; BAA89289.1; JOINED.
 DR EMBL; AB024503; BAA89289.1; JOINED.
 DR EMBL; AB024504; BAA89289.1; JOINED.
 DR EMBL; AB024505; BAA89289.1; JOINED.
 DR EMBL; AB024506; BAA89289.1; JOINED.
 DR EMBL; AF332078; AAK56107.1; -.
 DR EMBL; AF332080; AAK56108.1; -.
 DR EMBL; AF332080; AAK56108.1; -.
 DR PIR; JCS289; JCS289.
 DR MGD; MGI:108563; Ptpnbl.
 DR GO; GO:0008580; F:cytoskeletal regulator activity; IMP.
 DR GO; GO:0045309; F:phosphoprotein amino acid binding; IPT.
 DR GO; GO:0005515; F:protein binding; IPT.
 DR GO; GO:0007015; P:actin filament organization; IMP.
 DR GO; GO:0006928; P:cell motility; IMP.
 DR GO; GO:0007160; P:cell-matrix adhesion; IMP.

Query Match 6.0%; Score 203.5; DB 1; Length 513;
 Best Local Similarity 22.3%; Pred. No. 7.9e-06;
 Matches 92; Conservative 62; Mismatches 156; Indels 103; Gaps 18;

OY 248 SKSWITFDLKNKVEVVKVTDOPKQMGKLPMLHLTPQALP-----QYAGSGNULTALE 302
 DB 21 SASCTCTGVTGKELKVTQPEKSVVAAGDSTVLNCTLSLTPVGPPIKRYRGVGGSRLLIT 80
 OY 303 AKTGKLMHGVNLVVMRATQLOKNLTCFVWGPTSPKMLSLKLENKAKVSKREKRVWYLN 362
 DB 81 SFTGHPFRVTNV---SDATKRN-----NMDFSIHISN-----VTP 113
 OY 363 PEAGMQLDSDGVLLSINIKVLPWTSTVEPESCKTKTCCPCAPELL--GGSEVF 420
 DB 114 EDAGTYVVKFKQK-----PSEBPT-----EQSGGGEVY 144
 OY 421 LF---PK---PKDTLMISRTPEVTCVVVDSH--EDPEVKFMYVDGVENHAKTKPR 471
 DB 145 VLAKSPPEVSGPADRGIPDQKVNFTC---KSHQFSRNITLTKFKDGGSLHHEITVN 200
 OY 472 EEOYNSTYRVVSVLTVLHQDLNGEKYKVNKALPAPIEKITSKAGQPREPOVYTLR 531
 DB 201 PSKKNVSYNISSTYRV---LNSMDVASKVYCEVAHITLDR--SPLAGIANLNSFIRVS 254
 OY 532 PS-----RDELTKQVSTLCVKGYPESDIWENESNQEPENNYKTTTPVL--DSDSFF 584
 DB 255 PTVKVTQOSPVTSMQVNLTCRAERFYPDLQLIMLENGVSRN--DTPKNTLTKNTDGTYN 312
 OY 585 LYSKLTVDKSRMOGQNVFSGVMHE--ALNHYT-----QKSLSLSPG 625
 DB 313 YTSFLVNVSSAHRVDVFTCVQKHQDQALITRNHTVLGLAASSDQGSQGTTPG 365

RESULT 59
 SHS1_BOVIN STANDARD; PRT; 506 AA.

AC 046631; 046632; 046633; 046634; 046635; 046636; 046637; 046638; 046639; 046640; 046641; 046642; 046643; 046644; 046645; 046646; 046647; 046648; 046649; 046650; 046651; 046652; 046653; 046654; 046655; 046656; 046657; 046658; 046659; 046660; 046661; 046662; 046663; 046664; 046665; 046666; 046667; 046668; 046669; 046670; 046671; 046672; 046673; 046674; 046675; 046676; 046677; 046678; 046679; 046680; 046681; 046682; 046683; 046684; 046685; 046686; 046687; 046688; 046689; 046690; 046691; 046692; 046693; 046694; 046695; 046696; 046697; 046698; 046699; 046700; 046701; 046702; 046703; 046704; 046705; 046706; 046707; 046708; 046709; 046710; 046711; 046712; 046713; 046714; 046715; 046716; 046717; 046718; 046719; 046720; 046721; 046722; 046723; 046724; 046725; 046726; 046727; 046728; 046729; 046730; 046731; 046732; 046733; 046734; 046735; 046736; 046737; 046738; 046739; 046740; 046741; 046742; 046743; 046744; 046745; 046746; 046747; 046748; 046749; 046750; 046751; 046752; 046753; 046754; 046755; 046756; 046757; 046758; 046759; 046760; 046761; 046762; 046763; 046764; 046765; 046766; 046767; 046768; 046769; 046770; 046771; 046772; 046773; 046774; 046775; 046776; 046777; 046778; 046779; 046780; 046781; 046782; 046783; 046784; 046785; 046786; 046787; 046788; 046789; 046790; 046791; 046792; 046793; 046794; 046795; 046796; 046797; 046798; 046799; 046800; 046801; 046802; 046803; 046804; 046805; 046806; 046807; 046808; 046809; 046810; 046811; 046812; 046813; 046814; 046815; 046816; 046817; 046818; 046819; 046820; 046821; 046822; 046823; 046824; 046825; 046826; 046827; 046828; 046829; 046830; 046831; 046832; 046833; 046834; 046835; 046836; 046837; 046838; 046839; 046840; 046841; 046842; 046843; 046844; 046845; 046846; 046847; 046848; 046849; 046850; 046851; 046852; 046853; 046854; 046855; 046856; 046857; 046858; 046859; 046860; 046861; 046862; 046863; 046864; 046865; 046866; 046867; 046868; 046869; 046870; 046871; 046872; 046873; 046874; 046875; 046876; 046877; 046878; 046879; 046880; 046881; 046882; 046883; 046884; 046885; 046886; 046887; 046888; 046889; 046890; 046891; 046892; 046893; 046894; 046895; 046896; 046897; 046898; 046899; 046900; 046901; 046902; 046903; 046904; 046905; 046906; 046907; 046908; 046909; 046910; 046911; 046912; 046913; 046914; 046915; 046916; 046917; 046918; 046919; 046920; 046921; 046922; 046923; 046924; 046925; 046926; 046927; 046928; 046929; 046930; 046931; 046932; 046933; 046934; 046935; 046936; 046937; 046938; 046939; 046940; 046941; 046942; 046943; 046944; 046945; 046946; 046947; 046948; 046949; 046950; 046951; 046952; 046953; 046954; 046955; 046956; 046957; 046958; 046959; 046960; 046961; 046962; 046963; 046964; 046965; 046966; 046967; 046968; 046969; 046970; 046971; 046972; 046973; 046974; 046975; 046976; 046977; 046978; 046979; 046980; 046981; 046982; 046983; 046984; 046985; 046986; 046987; 046988; 046989; 046990; 046991; 046992; 046993; 046994; 046995; 046996; 046997; 046998; 046999; 047000; 047001; 047002; 047003; 047004; 047005; 047006; 047007; 047008; 047009; 047010; 047011; 047012; 047013; 047014; 047015; 047016; 047017; 047018; 047019; 047020; 047021; 047022; 047023; 047024; 047025; 047026; 047027; 047028; 047029; 047030; 047031; 047032; 047033; 047034; 047035; 047036; 047037; 047038; 047039; 047040; 047041; 047042; 047043; 047044; 047045; 047046; 047047; 047048; 047049; 047050; 047051; 047052; 047053; 047054; 047055; 047056; 047057; 047058; 047059; 047060; 047061; 047062; 047063; 047064; 047065; 047066; 047067; 047068; 047069; 047070; 047071; 047072; 047073; 047074; 047075; 047076; 047077; 047078; 047079; 047080; 047081; 047082; 047083; 047084; 047085; 047086; 047087; 047088; 047089; 047090; 047091; 047092; 047093; 047094; 047095; 047096; 047097; 047098; 047099; 047100; 047101; 047102; 047103; 047104; 047105; 047106; 047107; 047108; 047109; 047110; 047111; 047112; 047113; 047114; 047115; 047116; 047117; 047118; 047119; 047120; 047121; 047122; 047123; 047124; 047125; 047126; 047127; 047128; 047129; 047130; 047131; 047132; 047133; 047134; 047135; 047136; 047137; 047138; 047139; 047140; 047141; 047142; 047143; 047144; 047145; 047146; 047147; 047148; 047149; 047150; 047151; 047152; 047153; 047154; 047155; 047156; 047157; 047158; 047159; 047160; 047161; 047162; 047163; 047164; 047165; 047166; 047167; 047168; 047169; 047170; 047171; 047172; 047173; 047174; 047175; 047176; 047177; 047178; 047179; 047180; 047181; 047182; 047183; 047184; 047185; 047186; 047187; 047188; 047189; 047190; 047191; 047192; 047193; 047194; 047195; 047196; 047197; 047198; 047199; 047200; 047201; 047202; 047203; 047204; 047205; 047206; 047207; 047208; 047209; 047210; 047211; 047212; 047213; 047214; 047215; 047216; 047217; 047218; 047219; 047220; 047221; 047222; 047223; 047224; 047225; 047226; 047227; 047228; 047229; 047230; 047231; 047232; 047233; 047234; 047235; 047236; 047237; 047238; 047239; 047240; 047241; 047242; 047243; 047244; 047245; 047246; 047247; 047248; 047249; 047250; 047251; 047252; 047253; 047254; 047255; 047256; 047257; 047258; 047259; 047260; 047261; 047262; 047263; 047264; 047265; 047266; 047267; 047268; 047269; 047270; 047271; 047272; 047273; 047274; 047275; 047276; 047277; 047278; 047279; 047280; 047281; 047282; 047283; 047284; 047285; 047286; 047287; 047288; 047289; 047290; 047291; 047292; 047293; 047294; 047295; 047296; 047297; 047298; 047299; 047300; 047301; 047302; 047303; 047304; 047305; 047306; 047307; 047308; 047309; 047310; 047311; 047312; 047313; 047314; 047315; 047316; 047317; 047318; 047319; 047320; 047321; 047322; 047323; 047324; 047325; 047326; 047327; 047328; 047329; 047330; 047331; 047332; 047333; 047334; 047335; 047336; 047337; 047338; 047339; 047340; 047341; 047342; 047343; 047344; 047345; 047346; 047347; 047348; 047349; 047350; 047351; 047352; 047353; 047354; 047355; 047356; 047357; 047358; 047359; 047360; 047361; 047362; 047363; 047364; 047365; 047366; 047367; 047368; 047369; 047370; 047371; 047372; 047373; 047374; 047375; 047376; 047377; 047378; 047379; 047380; 047381; 047382; 047383; 047384; 047385; 047386; 047387; 047388; 047389; 047390; 047391; 047392; 047393; 047394; 047395; 047396; 047397; 047398; 047399; 047400; 047401; 047402; 047403; 047404; 047405; 047406; 047407; 047408; 047409; 047410; 047411; 047412; 047413; 047414; 047415; 047416; 047417; 047418; 047419; 047420; 047421; 047422; 047423; 047424; 047425; 047426; 047427; 047428; 047429; 047430; 047431; 047432; 047433; 047434; 047435; 047436; 047437; 047438; 047439; 047440; 047441; 047442; 047443; 047444; 047445; 047446; 047447; 047448; 047449; 047450; 047451; 047452; 047453; 047454; 047455; 047456; 047457; 047458; 047459; 047460; 047461; 047462; 047463; 047464; 047465; 047466; 047467; 047468; 047469; 047470; 047471; 047472; 047473; 047474; 047475; 047476; 047477; 047478; 047479; 047480; 047481; 047482; 047483; 047484; 047485; 047486; 047487; 047488; 047489; 047490; 047491; 047492; 047493; 047494; 047495; 047496; 047497; 047498; 047499; 047500; 047501; 047502; 047503; 047504; 047505; 047506; 047507; 047508; 047509; 047510; 047511; 047512; 047513; 047514; 047515; 047516; 047517; 047518; 047519; 047520; 047521; 047522; 047523; 047524; 047525; 047526; 047527; 047528; 047529; 047530; 047531; 047532; 047533; 047534; 047535; 047536; 047537; 047538; 047539; 047540; 047541; 047542; 047543; 047544; 047545; 047546; 047547; 047548; 047549; 047550; 047551; 047552; 047553; 047554; 047555; 047556; 047557; 047558; 047559; 047560; 047561; 047562; 047563; 047564; 047565; 047566; 047567; 047568; 047569; 047570; 047571; 047572; 047573; 047574; 047575; 047576; 047577; 047578; 047579; 047580; 047581; 047582; 047583; 047584; 047585; 047586; 047587; 047588; 047589; 047590; 047591; 047592; 047593; 047594; 047595; 047596; 047597; 047598; 047599; 047600; 047601; 047602; 047603; 047604; 047605; 047606; 047607; 047608; 047609; 047610; 047611; 047612; 047613; 047614; 047615; 047616; 047617; 047618; 047619; 047620; 047621; 047622; 047623; 047624; 047625; 047626; 047627; 047628; 047629; 047630; 047631; 047632; 047633; 047634; 047635; 047636; 047637; 047638; 047639; 047640; 047641; 047642; 047643; 047644; 047645; 047646; 047647; 047648; 047649; 047650; 047651; 047652; 047653; 047654; 047655; 047656; 047657; 047658; 047659; 047660; 047661; 047662; 047663; 047664; 047665; 047666; 047667; 047668; 047669; 047670; 047671; 047672; 047673; 047674; 047675; 047676; 047677; 047678; 047679; 047680; 047681; 047682; 047683; 047684; 047685; 047686; 047687; 047688; 047689; 047690; 047691; 047692; 047693; 047694; 047695; 047696; 047697; 047698; 047699; 047700; 047701; 047702; 047703; 047704; 047705; 047706; 047707; 047708; 047709; 047710; 047711; 047712; 047713; 047714; 047715; 047716; 047717; 047718; 047719; 047720; 047721; 047722; 047723; 047724; 047725; 047726; 047727; 047728; 047729; 047730; 047731; 047732; 047733; 047734; 047735; 047736; 047737; 047738; 047739; 047740; 047741; 047742; 047743; 047744; 047745; 047746; 047747; 047748; 047749; 047750; 047751; 047752; 047753; 047754; 047755; 047756; 047757; 047758; 047759; 047760; 047761; 047762; 047763; 047764; 047765; 047766; 047767; 047768; 047769; 047770; 047771; 047772; 047773; 047774; 047775; 047776; 047777; 047778; 047779; 047780; 047781; 047782; 047783; 047784; 047785; 047786; 047787; 047788; 047789; 047790; 047791; 047792; 047793; 047794; 047795; 047796; 047797; 047798; 047799; 047800; 047801; 047802; 047803; 047804; 047805; 047806; 047807; 047808; 047809; 047810; 047811; 047812; 047813; 047814; 047815; 047816; 047817; 047818; 047819; 047820; 047821; 047822; 047823; 047824; 047825; 047826; 047827; 047828; 047829; 047830; 047831; 047832; 047833; 047834; 047835; 047836; 047837; 047838; 047839; 047840; 047841; 047842; 047843; 047844; 047845; 047846; 047847; 047848; 047849; 047850; 047851; 047852; 047853; 047854; 047855; 047856; 047857; 047858; 047859; 047860; 047861; 047862; 047863; 047864; 047865; 047866; 047867; 047868; 047869; 047870; 047871; 047872; 047873; 047874; 047875; 047876; 047877; 047878; 047879; 047880; 047881; 047882; 047883; 047884; 047885; 047886; 047887; 047888; 047889; 047890; 047891; 047892; 047893; 047894; 047895; 047896; 047897; 047898; 047899; 047900; 047901; 047902; 047903; 047904; 047905; 047906; 047907; 047908; 047909; 047910; 047911; 047912; 047913; 047914; 047915; 047916; 047917; 047918; 047919; 047920; 047921; 047922; 047923; 047924; 047925; 047926; 047927; 047928; 047929; 047930; 047931; 047932; 047933; 047934; 047935; 047936; 047937; 047938; 047939; 047940; 047941; 047942; 047943; 047944; 047945; 047946; 047947; 047948; 047949; 047950; 047951; 047952; 047953; 047954; 047955; 047956; 047957; 047958; 047959; 047960; 047961; 047962; 047963; 047964; 047965; 047966; 047967; 047968; 047969; 047970; 047971; 047972; 047973; 047974; 047975; 047976; 047977; 047978; 047979; 04798

FT			(POTENTIAL).	
FT	MOD_RES	455	455	(PHOSPHORYLATION (BY TYR-KINASES)
FT	MOD_RES	472	472	(POTENTIAL).
FT	MOD_RES	498	498	(PHOSPHORYLATION (BY TYR-KINASES) (BY SIMILARITY)).
FT	MOD_RES	498	498	(PHOSPHORYLATION (BY TYR-KINASES) (BY SIMILARITY)).
FT	CARBOHYD	92	92	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	167	167	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	179	179	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	204	204	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	210	210	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	246	246	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	270	270	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	292	292	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	311	311	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	319	319	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	344	344	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	365	365	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	368	368	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT	23	23	T -> S.
FT	VARIANT	28	28	T -> A.
FT	VARIANT	61	61	S -> L.
FT	VARIANT	70	70	K -> R.
FT	VARIANT	120	120	Y -> H.
FT	VARIANT	125	125	R -> Q.
FT	VARIANT	127	127	E -> G.
FT	VARIANT	129	129	M -> H.
FT	VARIANT	132	132	R -> V.
FT	VARIANT	145	145	S -> N.
FT	VARIANT	153	153	L -> V.
FT	VARIANT	203	203	N -> D.
FT	VARIANT	261	261	G -> R.
FT	VARIANT	302	302	F -> L.
FT	VARIANT	316	316	F -> L.
FT	VARIANT	337	337	G -> R.
FT	VARIANT	367	367	S -> N.
FT	VARIANT	422	422	Q -> L.
FT	VARIANT	429	429	I -> F.
FT	VARIANT	433	433	D -> E.
SQL	SEQUENCE	506 AA;	55093 MW;	6B7E310677FC9CB CRC64;
<hr/>				
Query Match 5.5%; Score 187.5; DB 1; Length 506; Best Local Similarity 20.9%; Pred. No. 8.3e-05;				
Matches 89; Conservative 70; Mismatches 168; Indels 99; Gaps 17;				
Oy	246	SSSKSMITFDLNKKEVSVKRVYQDPDKLQMGGKLPLHLTLPOLP-----QYAGSGNLITA	300	
		:::::	:::::	
Db	20	AASNAM-TGTADGDELQVIQPERSVSAAGETATLHCTVLSIPVGPIKMFKGTA----	73	
Oy	301	LEAKTGK-LHOENVLVVMRATQL-----QKNLTCEWGPSTPKMLSLTLENKAEKVSR	354	
		:::::	:::::	
Db	74	----PGREFIYSQKAPPPRYTVNVS DATGRN-----NMDFSIIRSN-----	110	
Oy	355	EKPVVVLNPEDAGMOCCLSDSGOVLLESNIKVLPTWSTFVEBKSCDKTH-TCPPCPAEL	413	
		:::::	:::::	
Db	111	-----ITPADAGVYC-----VKFRKEERGDMEFKSGGFHTLVSAKRSPV	152	
Oy	414	LGPSVFLEPRPKKOTLMISRTPEVTCCVVVDVSH--EDEPVKFNNYDVGVENNAKTFR	471	
		:::::	:::::	
Db	153	LSGPFV-----RAFPEDVTNETCTSHSFSPRNISLWKFMKGNELISAQTSVD	199	
Oy	472	EEOVNSTYRVVSVLTVLHQDMNLNGEKYCKVENKALPA--PIEKTIISKAKQPREPOVYT	529	
Db	200	PEDNNVVISINSTTKVLLATGDSVHSQCIVEAHVTLGGPFL-----RGTAANSETIR	252	
Oy	530	LPSPDELT-----KNQVSLTCLKGYPSDIADVEMESNGOPENNKYTPRPVLDGSFF	584	
		:::::	:::::	
Db	253	VPPTL-EITGSPSAGNQVNVTCQMKFYPRHLQTLWELEGNMSRTREASVFEYNKDGTEN	311	
Oy	585	LYSKLTIVDKSRMQOGNVPSCSWMEALHNHYOKSLSLSPGLQDLBTCAADGEL----	640	
		:::::	:::::	
Db	312	QTSMELVNSSARRAEVNLTCQVEHNG-----QPVAVSKNHTLEVAPAPOKDODTQTPBN	365	

QY	641	DLMTT	646
DB	366	DSNMTS	371
RESULT	60		
VCAM1	RAT		
ID	VCAM1	RAT	STANDARD; PRT; 739 AA.
AC	P29534;		
DT	01-APR-1993	(Rel. 25, Created)	
DT	01-APR-1993	(Rel. 25, Last sequence update)	
DT	10-OCT-2003	(Rel. 42, Last annotation update)	
DE	Vascular cell adhesion protein 1 precursor (V-CAM 1).		
GN	VCAM1 OR VCAM-1.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.		
OX	NCBI_TaxId=10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Lung;		
RX	MEDLINE=92181437; PubMed=1371918;		
RA	Hession C., Moy P., Tizard R., Chisholm P., Williams C., Wysek M.,		
RA	Burdick L., Maye K., Kincaid P., Lobb R.;		
RT	"Cloning of murine and rat vascular cell adhesion molecule-1.";		
RL	Biochem Biophys Res Commun. 183:163-169(1992).		
CC	-1- FUNCTION: IMPORTANT IN CELL-CELL RECOGNITION. APPEARS TO FUNCTION		
CC	IN LEUKOCYTE-ENDOTHELIAL CELL ADHESION. INTERACTS WITH THE BETA-1		
CC	INTEGRIN VL44 ON LEUKOCYTES, AND MEDIATES BOTH ADHESION AND SIGNAL		
CC	TRANSDUCTION. THE VCAM1/VL44 INTERACTION MAY PLAY A		
CC	PATHOPHYSIOLOGIC ROLE BOTH IN IMMUNE RESPONSES AND IN LEUKOCYTE		
CC	EMIGRATION TO SITES OF INFLAMMATION.		
CC	-1- SUBCELLULAR LOCATION: Type 1 membrane protein.		
CC	-1- TISSUE SPECIFICITY: Expressed on inflamed vascular endothelium, as		
CC	well as on macrophage-like and dendritic cell types in both normal		
CC	and inflamed tissue.		
CC	-1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; M84488; AAA42332.1; -.		
DR	PIR; J50675; J50675.		
DR	HSSP; P19930; P19930.		
DR	InterPro; IPR003987; ICAM VCAM-1.		
DR	InterPro; IPR007110; Ig-like.		
DR	InterPro; IPR003598; Ig_c2.		
DR	InterPro; IPR003989; VCAM-1.		
DR	PIFAM; PF00047; Ig; 5.		
DR	PRINTS; PR01472; ICAMVCAM1.		
DR	PRINTS; PR01474; VCAM1.		
DR	SMART; SM00408; IGC2; 3.		
DR	PROSITE; PS50835; IG_LIKE; 5.		
KW	Immunoglobulin domain; Glycoprotein; Cell adhesion; Transmembrane;		
RV	Repeat; Signal.		
FT	SIGNAL	1	24
FT	CHAIN	25	739
FT	DOMAIN	25	698
FT	TRANSMEM	699	720
FT	DOMAIN	721	739
FT	DOMAIN	25	111
FT	DOMAIN	119	212
FT	DOMAIN	223	309
FT	DOMAIN	312	397
FT	DOMAIN	408	506
FT	DOMAIN	514	595
FT	DOMAIN	601	682
FT	PROBABLE. VASCULAR CELL ADHESION PROTEIN 1.		
FT	EXTRACELLULAR (POTENTIAL).		
FT	CYTOPLASMIC (POTENTIAL).		
FT	IG-LIKE C2-TYPE 1.		
FT	IG-LIKE C2-TYPE 2.		
FT	IG-LIKE C2-TYPE 3.		
FT	IG-LIKE C2-TYPE 4.		
FT	IG-LIKE C2-TYPE 5.		
FT	IG-LIKE C2-TYPE 6.		
FT	IG-LIKE C2-TYPE 7.		

```

FT DISULFID 47 95 BY SIMILARITY.
FT DISULFID 52 99 BY SIMILARITY.
FT DISULFID 137 195 BY SIMILARITY.
FT CARBOHYD 273 273 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 531 531 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 561 561 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 650 650 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 739 AA; 81246 MW; 5C6085A1A1B100C CRC64;

Query Match 5.3%; Score 182.5; DB 1; Length 739;
Beet Local Similarity 20.8%; Pred.No.0.00028;
Matches 124; Conservative 87; Mismatches 195; Indels 191; Gaps 28;

OY 33 KGDYELCTASOKSIOFHWKNSNQIKLGNQSGFLFKGSKLNDRADRSRLMDQGNF 92
DB EGAANTWCASGCLPAPLFIWMSK-----LNGVQL-----SNA 274
OY 93 PLIKNKLESDTYICE---VEDQKEVOLV-----FGLTNSDTHLLQGSITLT 142
DB 275 TLTLIAMEDESGIYCEGVNLVGRDKTEVELIQEKFPYVDISPGSQAQVGDVYLT 334
OY 143 LESPPGSSPVOCRRPRGNKINOG-----GKTLVSQLEQDSTWTCTVLQNKRYEF 195
DB 335 CAAVGCDSPFSWRTQUTDSPLNGEVDRGATSTLTLSFVGVEDEHSYLCYTCORRKLK 394
OY 196 KIDIVLAFQKASIVYKKEGQVEFPLAFTVFKLTGSGELMWQABASSKSWITFD 255
DB 395 TIQEVVYF-----PEDEITISGLVH----- 417
OY 256 LKNEVSVKRYVTQDPKLOMKKPLHLTLPOALPOYAGSGLTALAKTKLQEVNLV 315
DB 418 -----GRPVYNCVPMVYFPD-----HLEIELL 441
OY 316 VMRATQLOKULTCEWGPPTSPKMLSLLENKEAVSREKRVWVLEAGMQLSDS 375
DB 442 KGETLLMKFRLREI-GTKS-----LETSLMT-----FIPTA-----EDT 477
OY 376 GOVLLESNIKVLPTWSTVEPEKSCDKTATCPCPAPPELLGSPVFLFP-PRKDTIMSR 434
DB 478 GKALV--CLAKLHSSQMSSEPRQROSTQTLVYNAFK---EPTIWSSPPEBESPN- 531
OY 435 TPVATCVVDVSHEDPEVKFMYVGVGVHNAKTKPREQYNSYTVRVSVLTVLHQMLN 494
DB 532 ---LTC---SSDGPPTPKILM---SRQLKNGELOPLSQ---NTL---LSFMATKMD--- 573
OY 495 GREYCKXSNKALPRLPIKITSK-----ARGQPREPVYTLPPSRDELTKQVSLTGLV 548
DB 574 SGIVYCEGINEA-----GISKSKVELLIQSSSKDIDL-TVPBSKYVEGDVTYISCTC 625
OY 549 KGFYPSDIAVEMESNGQPENNYKTPPYLDS-DGSFFLYSKLTVDKSRWQGNVFS 604
DB 626 -GSVEPEIIT-----LKKKAKTODMWLKSVNGSY-----TIKQALQDAGYIEC 668

RESULT 61
PGBM_HUMAN STANDARD; PRT; 4391 AA.
AC P98150; 016287; 09H3V5;
DT 01-OCT-1996 (Rel. 34, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Basement membrane-specific heparan sulfate proteoglycan core
DE protein precursor (HSPG) (perlecan) (PLC).
GN HSPG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9211294; PubMed=1730768;
RA Kallunki P., Tryggvason K.;

```

```

RT RT "Human basement membrane heparan sulfate proteoglycan core protein: a
RT RT 467-kD protein containing multiple domains resembling elements of the
RT RT low density lipoprotein receptor, laminin, neural cell adhesion
RT RT molecules, and epidermal growth factor.";
RT RT J. Cell Biol. 116:559-571(1992).
RL [2]
RN RN SEQUENCE FROM N.A.
RP RP TISSUE=Colon, and Skin;
RC RC MEDLINE=92235084; PubMed=1569102;
RA RA Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
RT RT "Primary structure of the human heparan sulfate proteoglycan from
RT RT basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
RT RT domains homologous to the low density lipoprotein receptor, laminin,
RT RT neural cell adhesion molecules, and epidermal growth factor.";
RT RT J. Biol. Chem. 267:8544-8557(1992).
RN [3]
RP RP SEQUENCE OF 22-4391 FROM N.A., AND VARIANT SISI TYR-1532.
RX RX MEDLINE=2053141; PubMed=1101850;
RA RA Nicole S., Davoine C.-S., Topaloglu H., Cattolico L., Barral D.,
RA Beigton P., Ben-Hamida C., Hammouda H., Cruaud C., White P.S.,
RA Samson D., Urtizberea J.A., Lehmann-Horn F., Weissenbach J.,
RA Hentati F., Fontaine B.;
RT RT "Perlecan, the major proteoglycan of basement membranes, is altered in
RT RT patients with Schwartz-Jampel syndrome (chondrodystrophic myotonia)."
RT RT Nat. Genet. 26:480-483(2000).
RN [4]
RP RP SEQUENCE OF 1016-1470 FROM N.A.
RC RC TISSUE=Colon;
RX RX MEDLINE=91365376; PubMed=1679749;
RA RA Dodge G.R., Kovalszky I., Chu M.L., Haessel J.R., McBride O.W.,
RA Yi H.F., Iozzo R.V.;
RT RT "Heparan sulfate proteoglycan of human colon: partial molecular
RT RT cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT RT short arm of human chromosome 1.";
RT RT Genomics 10:673-680(1991).
RN [5]
RP RP SEQUENCE OF 890-1396 FROM N.A.
RC RC TISSUE=Fibroblastoma;
RX RX MEDLINE=92120660; PubMed=1685141;
RA RA Kallunki P., Eddy R.L., Byers M.G., Keestila M., Shows T.B.,
RA Tryggvason K.;
RT RT "Cloning of human heparan sulfate proteoglycan core protein,
RT RT assignment of the gene (HSPG2) to 1p36.1--p35 and identification of
RT RT a BamHI restriction fragment length polymorphism.";
RT RT Genomics 11:389-396(1991).
RN [6]
RP RP SEQUENCE OF 1-21 FROM N.A.
RX RX MEDLINE=94052171; PubMed=8234307;
RA RA Cohen I.R., Graessel S., Murdoch A.D., Iozzo R.V.;
RT RT "Structural characterization of the complete human perlecan gene and
RT RT its promoter.";
RT RT Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
RN [7]
RP RP CARBOHYDRATE-LINKAGE SITE ASN-2121.
RX RX MEDLINE=22660472; PubMed=12754519;
RA RA Zhang H., Li X.-J., Martin D.B., Aebersold R.;
RT RT "Identification and quantification of N-linked glycoproteins using
RT RT hydrazide chemistry, stable isotope labeling and mass spectrometry.";
RT RT Nat. Biotechnol. 21:660-666(2003).
RN [8]
RP RP FUNCTION: This protein is an integral component of basement
RP RP membranes. It is responsible for the fixed negative electrostatic
RP RP charge and is involved in the charge-selective ultrafiltration
RP RP properties. It serves as an attachment substrate for cells.
RN [9]
RP RP SUBUNIT: Purified perlecan has a strong tendency to aggregate in
RP RP dimers or stellate structures. It interacts with other basement
RP RP membrane components such as laminin, prolargin and collagen type
RP RP IV.
RN [10]
RP RP SUBCELLULAR LOCATION: Extracellular.
RN [11]
RP RP TISSUE SPECIFICITY: Found in the basement membranes.
RN [12]
RP RP PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
RN [13]
RP RP AND O-LINKED OLIGOSACCHARIDES.
RN [14]
RP RP DISEASE: Defects in HSPG2 are the cause of Schwartz-Jampel
RN [15]
RP RP syndrome (SUS1) [MIM:255800]; a rare autosomal recessive disorder

```



```

Db      2849  VPPGAAHQVTHWKKGNLPAR-HQVHGRPLRLNVSADSGEYSCQVTGSSGT-LEASV 2906
Oy      385  KVLPTWSTPVEPKSCDKHTHCP-PCPAPELLGGPVEFLFPPPKKDTLMISRTPEVTCYV 443
Db      2907  LV-----TIETSS-----PGFLPAP-----GLAQPIYIESSSHVTEGQTLINCYVP 2949
Oy      444  DVSHEDPEVKFNVDVGVENHNAKTRPPEEQYNSTYRVVSVLTVLHQMNGKEYCKVVS 503
Db      2950  GOAH-----AQVTHWYKRG-----GSLPARHQTHGSLRLHLVSPA-----DSGEYVCRAA 2994
Oy      504  NKALP-----APIEKTISKAKGQP-----REPQVYTLPPSRBELTNQVSLTCLV-KGYPSPDI 556
Db      2995  SGPGGEQGEASTVTYVPPSEGSRYLRSPVISTDPPSSYVQQGDASFKCLIHDAAP--I 3052
Oy      557  AVEMESNGQP-ENNYKTPPVLDSDGSFFLYSKLTIVDKSRMQQGVFSC-----S 605
Db      3053  SLEWTKRQDELEDNVHISP-----NGSI-----ITIVGTRPSNHCTYCVASNAVGAQS 3102
Oy      606  VMHEALHNHYTKSLISLP-----GLQDETCAEQDELDGLWT 645
Db      3103  VVNLISVHGPPPTVSVLPBGPVWVKVKAATLBCVSAGBPRSSARWT 3147

```

```

RESULT 62
LAC1_MOUSE
ID      LAC1_MOUSE  STANDARD;  PRT;  105 AA.
AC      P01843;
DT      21-JUL-1966 (Rel. 01, Created)
DT      21-JUL-1966 (Rel. 01, Last sequence update)
DT      15-MAR-2004 (Rel. 43, Last annotation update)
DE      Ig lambda-1 chain C region.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxId=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      MEDLINE=83014953; PubMed=6812053;
RA      Seising E., Miller J., Wilson R., Storb U.;
RT      "Evolution of mouse immunoglobulin lambda genes.";
RL      Proc. Natl. Acad. Sci. U.S.A. 79:4681-4685(1982).
RN      [2]
RP      SEQUENCE FROM N.A. (MOPC 315).
RA      MEDLINE=81148606; PubMed=6259534;
RA      Bothwell A.L.M., Paekind M., Schwartz R.C., Sonenshein G.E.,
RT      "Dual expression of lambda genes in the MOPC-315 plasmacytoma.";
RL      Nature 290:65-67(1981).
RN      [3]
RP      SEQUENCE FROM N.A. (S43).
RA      MEDLINE=82220143; PubMed=6283385;
RA      Bothwell A.L.M., Paekind M., Reth M., Imanishi-Kari T., Rajewsky K.,
RT      "Somatic variants of murine immunoglobulin lambda light chains.";
RL      Nature 298:380-382(1982).
RN      [4]
RP      SEQUENCE (MYELOMA PROTEIN MOPC 104E).
RA      MEDLINE=71107854; PubMed=5276767;
RA      Appella E.;
RT      "Amino acid sequences of two mouse immunoglobulin lambda chains.";
RL      Proc. Natl. Acad. Sci. U.S.A. 68:590-594(1971).
CC      -1- MISCELLANEOUS: The MOPC 315 cell line produces 2 light chains, 1
CC      normal lambda-2 chain and 1 abnormal lambda-1 chain that is
CC      missing a large part of the V region. The C region sequence (shown
CC      here) appears completely normal.
CC      -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; J00582; AAA51636.1; -.
DR      EMBL; J00587; AAB59672.1; -.
DR      PIR; A93922; LIMS.
DR      PDB; 1JNH; 06-FEB-02.
DR      InterPro; IPR007110; Ig-like.
DR      InterPro; IPR003597; Ig_c1.
DR      InterPro; IPR003006; Ig_MHC.
DR      Pfam; PF00047; Ig; 1.
DR      SMART; SM00407; Igcl; 1.
DR      PROSITE; PS50835; IG_LIKE; 1.
DR      PROSITE; PS00290; IG_MHC; 1.
KW      Immunoglobulin domain; Immunoglobulin C region; 3D-structure.
FT      NON TER 1 1
FT      DOMAIN 1 1
FT      DISULFID 27 100 IG-LIKE.
FT      DISULFID 104 104 INTERCHAIN (WITH HEAVY CHAIN).
FT      CONFLICT 19 20 ET -> TE (IN REF. 4).
FT      CONFLICT 56 56 Q -> E (IN REF. 4).
FT      CONFLICT 75 75 MISSING (IN REF. 4).
FT      CONFLICT 81 82 HS -> SH (IN REF. 4).
FT      CONFLICT 85 85 S -> SS (IN REF. 4).
FT      CONFLICT 96 96 E -> Q (IN REF. 4).
SQ      SEQUENCE 105 AA; 11575 MW; A89F2B09BCFCA018 CRC64;

```

```

Query Match 5.2%; Score 177.5; DB 1; Length 105;
Best Local Similarity 35.9%; Pred. No. 4,6e-05;
Matches 37; Conservative 18; Mismatches 43; Indels 5; Gaps 3;

```

```

Oy      521  QPR-EPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQP-ENNYKTPPVLD 578
Db      1  QPKSSPVTLFPPSESELETKATLVCTITDFYGVATVWVKVDTGTQGMETOPSKQ 60
Oy      579  SDGSFFLYSKLTVDKSRMQQGVFSCVMHEALHNHYTKSLIS 621
Db      61  SNKRYMASSYTLTAPARAMERRHSSYSCQVTHE--GHTVERKLS 100

```

```

RESULT 63
UN89_CABEL
ID      UN89_CABEL  STANDARD;  PRT;  6632 AA.
AC      001761; Q17362;
DT      10-OCT-2003 (Rel. 42, Created)
DT      10-OCT-2003 (Rel. 42, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN      UNC-89 OR C09D1.1.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC      Rhabditidae; Pelodierinae; Caenorhabditis.
OX      NCBI_TaxId=6239;
RN      [1]
RP      SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC      STRAIN=Bristol N2;
RX      MEDLINE=96180278; PubMed=8603916;
RA      Beilan G.M., Tinley T.L., Tang X., Borodovsky M.;
RT      "The Caenorhabditis elegans gene unc-89, required for muscle M-line
RT      assembly, encodes a giant modular protein composed of Ig and signal
RT      transduction domains.";
RL      J. Cell Biol. 132:835-848(1996).
RN      [2]
RP      SEQUENCE FROM N.A.
RA      STRAIN=Bristol N2;
RC      Du Z., Le T.T., Wilson R.;
RL      Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      REVISIONS.
RA      Waterston R.;
RL      Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC      -1- FUNCTION: Structural component of the muscle M-line. Myofibrament
CC      lattice assembly begins with positional cues laid down in the
CC      basement membrane and muscle cell membrane. UNC-89 responds to

```

these signals, localizes, and then participates in assembling an M-line.

- TISSUE SPECIFICITY: Localizes to the middle of A-bands.

- SIMILARITY: Contains 1 DBL-homology (DH) domain.

- SIMILARITY: Contains 1 fibronectin type III domain.

- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.

- SIMILARITY: Contains 1 PH domain.

- SIMILARITY: Contains 5 RCHD domains.

- SIMILARITY: Contains 1 SH3 domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).

EMBL, U3058; AAB00542.1; -

EMBL, AF003131; AAB54132.2; -

PDB: 1FHO; 20-DEC-00

WormPep: C09D1.1; CB30426.

Interp: IPR008957; FN_III-like.

Interp: IPR003961; FN_III.

Interp: IPR007110; IG-like.

Interp: IPR003598; IG_c2.

Interp: IPR003006; IG_MHC.

Interp: IPR001849; PH.

Interp: IPR007850; RCHD.

Interp: IPR000219; RhGEF.

DR pfam: PF00041; fn3; 1.

DR pfam: PF00047; ig; 47.

DR pfam: PF00169; PH; 1.

DR pfam: PF05177; RCHD; 5.

DR pfam: PF00621; RhGEF; 1.

DR pfam: PF00018; SH3; 1.

DR SMART; SM00408; IGC2; 23.

DR SMART; SM00325; RhGEF; 1.

DR SMART; SM00326; SH3; 1.

DR PROSITE; PS50010; DH_2; 1.

DR PROSITE; PS50835; IG-like; 49.

DR PROSITE; PS50003; PH DOMAIN; 1.

DR PROSITE; PS50002; SH3; 1.

KM Muscle protein, immunoglobulin domain; Repeat; SH3 domain; 3D-structure.

KM 3D-structure.

FT DOMAIN 63 127 SH3.

FT DOMAIN 152 330 DH.

FT DOMAIN 342 498 PH.

FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.

FT DOMAIN 648 736 IG-LIKE C2-TYPE 2.

FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.

FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.

FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.

FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.

FT DOMAIN 1272 1315 THR-RICH.

FT DOMAIN 1375 1475 RCHD 1.

FT DOMAIN 1479 1585 RCHD 2.

FT DOMAIN 1597 1695 RCHD 3.

FT DOMAIN 1700 1799 RCHD 4.

FT DOMAIN 1800 1860 RCHD 5.

FT DOMAIN 1860 1960 RCHD 5.

FT DOMAIN 1960 2067 IG-LIKE C2-TYPE 7.

FT DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.

FT DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.

FT DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.

FT DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.

FT DOMAIN 2463 2564 IG-LIKE C2-TYPE 12.

FT DOMAIN 2563 2651 IG-LIKE C2-TYPE 13.

FT DOMAIN 2657 2746 IG-LIKE C2-TYPE 14.

FT DOMAIN 2754 2858 IG-LIKE C2-TYPE 15.

FT DOMAIN 2887 2980 IG-LIKE C2-TYPE 16.

FT DOMAIN 2994 3081 IG-LIKE C2-TYPE 17.

FT DOMAIN 3087 3183 IG-LIKE C2-TYPE 18.

FT DOMAIN 3189 3280 IG-LIKE C2-TYPE 19.

FT DOMAIN 3286 3376 IG-LIKE C2-TYPE 20.

FT DOMAIN 3384 3469 IG-LIKE C2-TYPE 21.

FT DOMAIN 3482 3572 IG-LIKE C2-TYPE 22.

FT DOMAIN 3580 3667 IG-LIKE C2-TYPE 23.

FT DOMAIN 3686 3777 IG-LIKE C2-TYPE 24.

FT DOMAIN 3817 3908 IG-LIKE C2-TYPE 25.

FT DOMAIN 3920 4009 IG-LIKE C2-TYPE 26.

FT DOMAIN 4018 4106 IG-LIKE C2-TYPE 27.

FT DOMAIN 4109 4201 IG-LIKE C2-TYPE 28.

FT DOMAIN 4212 4297 IG-LIKE C2-TYPE 29.

FT DOMAIN 4302 4387 IG-LIKE C2-TYPE 30.

FT DOMAIN 4400 4485 IG-LIKE C2-TYPE 31.

FT DOMAIN 4489 4580 IG-LIKE C2-TYPE 32.

FT DOMAIN 4588 4678 IG-LIKE C2-TYPE 33.

FT DOMAIN 4681 4771 IG-LIKE C2-TYPE 34.

FT DOMAIN 4771 4861 IG-LIKE C2-TYPE 35.

FT DOMAIN 4861 4951 IG-LIKE C2-TYPE 36.

FT DOMAIN 4965 5057 IG-LIKE C2-TYPE 37.

FT DOMAIN 5067 5160 IG-LIKE C2-TYPE 38.

FT DOMAIN 5171 5260 IG-LIKE C2-TYPE 39.

FT DOMAIN 5277 5366 IG-LIKE C2-TYPE 40.

FT DOMAIN 5383 5472 IG-LIKE C2-TYPE 41.

FT DOMAIN 5487 5578 IG-LIKE C2-TYPE 42.

FT DOMAIN 5595 5685 IG-LIKE C2-TYPE 43.

FT DOMAIN 5701 5790 IG-LIKE C2-TYPE 44.

FT DOMAIN 5815 5904 IG-LIKE C2-TYPE 45.

FT DOMAIN 5925 6014 IG-LIKE C2-TYPE 46.

FT DOMAIN 6038 6130 IG-LIKE C2-TYPE 47.

FT DOMAIN 6150 6239 IG-LIKE C2-TYPE 48.

FT DOMAIN 6275 6368 FIBRONECTIN TYPE-III.

FT DOMAIN 6413 6502 IG-LIKE C2-TYPE 49.

FT DOMAIN 6507 6596 POTENTIAL.

FT DISULFID 568 621 POTENTIAL.

FT DISULFID 2908 2975 POTENTIAL.

FT DISULFID 3015 3065 POTENTIAL.

FT DISULFID 3707 3759 POTENTIAL.

FT DISULFID 3826 3890 POTENTIAL.

FT DISULFID 5092 5157 POTENTIAL.

FT DISULFID 5288 5350 POTENTIAL.

FT DISULFID 5508 5560 POTENTIAL.

FT DISULFID 5616 5669 POTENTIAL.

FT DISULFID 5722 5764 POTENTIAL.

FT DISULFID 5836 5901 POTENTIAL.

FT DISULFID 5946 5998 POTENTIAL.

FT DISULFID 6036 6171 POTENTIAL.

FT DISULFID 6421 6486 POTENTIAL.

FT DISULFID 2137 2137 AKA -> P (IN REF. 1).

FT CONFLICT 2245 2247 AKA -> PKP (IN REF. 1).

FT CONFLICT 2258 2258 E -> G (IN REF. 1).

FT CONFLICT 2284 2284 M -> I (IN REF. 1).

FT CONFLICT 2287 2287 A -> G (IN REF. 1).

FT CONFLICT 3531 3531 DAGEV -> RRRRI (IN REF. 1).

FT CONFLICT 3884 3888 A -> V (IN REF. 1).

FT CONFLICT 3929 3929 A -> P (IN REF. 1).

FT CONFLICT 5134 5134 T -> S (IN REF. 1).

FT CONFLICT 5145 5145 G -> A (IN REF. 1).

FT CONFLICT 5165 5185 K -> N (IN REF. 1).

FT CONFLICT 5199 5199 L -> F (IN REF. 1).

FT CONFLICT 5202 5202 F -> L (IN REF. 1).

FT CONFLICT 5213 5213 F -> G (IN REF. 1).

FT CONFLICT 6178 6178 K -> E (IN REF. 1).

FT CONFLICT 6268 6268 K -> E (IN REF. 1).

SQ SEQUENCE 6632 AA; 731665 MW; 2623EDD62960E89 CRC64;

Query Match 5.1%; Score 174.5; DB 1; Length 6632;

Best Local Similarity 22.2%; Pred. No. 0.017;

Matches 127; Conservative 75; Mismatches 200; Indels 171; Gaps 27;

QY 34 GDTVELTCTAQCKSIQFHW-KNSNQIKILNQGSFLTKGSKLNDRADRSRLMDQGNF 92

DB 3934 GETAVLTBCKIGSKGKPSVYKWKNGEELK-----PS---DRVKIEN--LDDSTQ 3976

QY 93 PLIINKLIEBDSDTYICGEVDQKEEVGLVFGFLTANSDDTHLLQGQSLTLTLSPGSGSPS 152

```

Db 3977 RLTVNADLDMDVEYRCASNEFGDV-----MSDVLTYKEPQVAFG 4019
Oy 153 VOCRSPRGKNIQGGKTLISVSOLEDSGTWTCTVLQONOKVEFKIDIVLAFOKASSIVY 212
Db 4020 F-----FELSAIQVETETAKFECKVSGTSPDKM-----FKQGTPL-- 4057
Oy 213 KKEGQVEFSPFLAFTVEKLTGSGELMWQABASSKSMITFDLKNKEVSKRYTQDKL 272
Db 4058 -KEDKRVHF-----ESTDDG-----TQRLVEDSKT 4082
Oy 273 QMGKTLPHLTLPLQALPQVAGSNLTALAEKTLGHQVNLVYM--RATLOKRLTCEV 330
Db 4083 D-----DQGNRIEVSNDAGVANSKVPITVVPSETLXIKKGLT-DV 4122
Oy 331 WGPSPKMLSLKLENKEAKV-----SKREKPVWVLANPE--AGMMQCLLSDSGQ 377
Db 4123 NVTQGTKLLSVEVGKPKTKYKMYGTETVSSQCTTKVQVTESEYKLEISAKMSDIGA 4182
Oy 378 VLESNIKVLPWTSPVEPKSCDKTHTCPCPAPELLGQSVFLPPPKDQTLMSRTP 437
Db 4183 YRV-----VLSTDSFSVESSA-----TVTVTKAAEKISLPSP-----KKGADQSVK 4225
Oy 438 VTCVVVDVSHEDPEVKFMYVDGVEVHNAKTKRREQYNSTRYVSULTVHQMWNKE 497
Db 4226 GTPPLVLEIEIGKPKDVKMYKNGDEIKDQKV---EDLNGKRY---LTIPDFQEKDVG 4278
Oy 498 YKCKSNKALPAPIEKTSKAK--GQRPPOVYT--LPPS---RDELTKQVSLTLCVGFY 552
Db 4279 YSVTPANBA--GEIF---SKAKNVSAKPEIVSGVLPFTVQGEYATNVAKVPVK-- 4331
Oy 553 PSDIAVENESNGQPENNYKTPPVLDSDGSFFL 585
Db 4332 -----VKWYKNGKEIPDAKTD--NGDGSYSL 4356

RESULT 64
LAC2_RAT STANDARD; PRT; 104 AA.
AC P20767;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 1g lambda-2 chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87305594; PubMed=3114047;
RA Steen M.L., Hellman L., Petersson U.;
RT "The immunoglobulin lambda locus in rat consists of two C lambda
RT genes and a single V lambda gene.";
RL Gene 55:75-84(11987).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M22521; AAA41420.1; ALT_INT.
DR HSPB; P01842; 2MCG.
DR InterPro; IPR007110; 1g-1like.
DR InterPro; IPR003597; 1g_c1.
DR InterPro; IPR003006; 1g_MHC.
DR Pfam; PF00047; 1g; 1.
DR SMART; SM00407; Tgcl; 1.
DR PROSITE; PS50835; 1g_LIKE; 1.

```

```

DR PROSITE; PS00290; 1g_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region.
FT NON TER 1
FT DOMAIN 6 99 IG-LIKE.
FT DISULFID 27 85
FT DISULFID 103 103 INTERCHAIN (WITH HEAVY CHAIN).
SQ SEQUENCE 104 AA; 11318 MW; F087906DE43F7276 CRC64;

Query Match 5.0%; Score 172; DB 1; Length 104;
Best Local Similarity 35.3%; Pred. No. 0.0001;
Matches 36; Conservative 17; Mismatches 45; Indels 4; Gaps 2;

Oy 521 QPRE-POVYTLPPSRDELTKQVSLTLCVKGFPSPDIAVENESNGQPENNYKTPPVLD 579
Db 1 QPKSTPLTVFPSPSTELQGNKATLVCLISDFPSDVEVAMKNGAPISQGVDTANPTKQ 60
Oy 580 DGSFELSKLTVDKSRWQGNVSCSVMEALNHNHYQSL 621
Db 61 GNKIYASSFLRLTAQWRNSRSTFCQVTHF--GNVTKSL 99

RESULT 65
PGBM_MOUSE STANDARD; PRT; 3707 AA.
AC Q05793;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Basement membrane-specific heparan sulfate proteoglycan core
DE protein precursor (HSPG) (Perlecan) (PLC).
GN HSPG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Melanoma;
RX MEDLINE=92078153; PubMed=1744087;
RA Noonan D.M., Horigan E.A., Ledbetter S.R., Vogel G., Sasaki M.,
RA Yamada Y., Hassell J.R.;
RT "The complete sequence of perlecan, a basement membrane heparan
RT sulfate proteoglycan, reveals extensive similarity with laminin A
RT chain, low density lipoprotein-receptor, and the neural cell adhesion
RT molecule.";
RL J. Biol. Chem. 266:22939-22947(1991).
RN [2]
RP SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=89034110; PubMed=2972708;
RA Noonan D.M., Horigan E.A., Ledbetter S.R., Vogel G., Sasaki M.,
RA Yamada Y., Hassell J.R.;
RT "Identification of cDNA clones encoding different domains of the
RT basement membrane heparan sulfate proteoglycan.";
RL J. Biol. Chem. 263:16379-16387(1988).
CC -1- FUNCTION: This protein is an integral component of basement
CC membranes. It is responsible for the fixed negative electrostatic
CC charge and is involved in the charge-selective ultrafiltration
CC properties. It serves as an attachment substrate for cells.
CC -1- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC dimers or stellate structures. It interacts with other basement
CC membrane components such as laminin, prolargin and collagen type
CC IV.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: Found in the basement membranes.
CC -1- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC AND O-LINKED OLIGOSACCHARIDES.
CC -1- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC -1- SIMILARITY: Contains 11 laminin EGF-like domains.
CC -1- SIMILARITY: Contains 3 laminin IV domains.
CC -1- SIMILARITY: Contains 15 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 3 laminin G-like domains.
CC -1- SIMILARITY: Contains 1 EGF-like domain.
CC -1- SIMILARITY: Contains 1 SEA domain.

```



```

FT DISULFID 1976 2021 BY SIMILARITY.
FT DISULFID 2073 2118 BY SIMILARITY.
FT DISULFID 2170 2215 BY SIMILARITY.
FT DISULFID 2268 2313 BY SIMILARITY.

Query Match 4.9%; Score 168.5; DB 1; Length 3707;
Best Local Similarity 17.7%; Pred. No. 0.019;
Matches 116; Conservative 94; Mismatches 251; Indels 193; Gaps 26;

QY 33 KGDVTEICTASQKSIQPHMNSNQIKLNGSGFLTKPSKULNDRADSRSLMDQNF 92
DB 1968 EGRVRLYCRAGVPSASITWKEG-----GSLPRHQHNSR----- 2005
QY 93 PLIINKLIEDSDTYIC-----EVEDQKEVOLLYFGLT-----ANSDTHLL 134
DB 2006 -LRHHMVSADSGEVYVCRRANNIDQERTSIMSVSPNSPPAPAPARIIESSSSVA 2064
QY 135 QGQSITLTLESPPGSSPSVQCRSPRGKNI-----QGKTLVSQLELODSTGTWTYQ 189
DB 2065 EGQTLDLNCVVPGHAAQVTHMK-RGSGLPTHQTHGSRRLRYQVSSADSGEVYCVSLSS 2123
QY 190 QKVEFKDIYVLARQKASSTIYKKEGQVSPFLATVTEKLTGSGELMWQAERASSK 249
DB 2124 SGPLASVAVST-----PAAANVH----- 2143
QY 250 SWITFDLKNKEVSXRVVTOQPKLQMGKLLPHLTLPQALPYAGSGNLTALAEKTKLH 309
DB 2144 -----IGVVPRIETSSSRVAGQTLDSLQVPP-----GQAH 2177
QY 310 QEVNLVNRATQLOKNTLCEVWGPRTSPKMLSLKLENKAVSKREKVPWVNLNPAQMK 369
DB 2178 AQV-----TWKKRGSLPAGHQVGHMLNLR-----VSPADSGEYS 2214
QY 370 CLLSQSGVLESNKIVLPTWSTPYERPKSCDHTHCPCRPAPPELLGSPVLFPPKPKDT 429
DB 2215 CQVTSSTGCT-LEASVLTITESEP-----STIPAP-----GLAQPVYIESSSSH 2257
QY 430 LMISRTPEVTCVVVDVSHEDPEVKFPMWVDGVEVNAATKREEQYNSTYVAVLTLH 489
DB 2258 LTBGGTVDLKKCVPEQAH-----AQVTWTKRG-----SLPRHQHNSLALYQLSPA-- 2306
QY 490 QDWLNGKEYKCKVSNKALP---APIETKISAKAQGP---RBPQVYLTTPSDELTQKQVS 543
DB 2307 ----DSGEYVCQVAGSSSHPEHSAFKLTVPSGSSFLRBPVSIIEPPSSTVQGGQDAS 2362
QY 544 LTCLV-KGFYPSDLAVEMESNQF-ENNYKTPPLDSDGSFLLX---SKLTVDKSRQO 598
DB 2363 FKCLIHGAMP--LKEVKKIRDQLELDVHISP-----NGSIITITVADGPATMEPTACVA 2415
QY 599 GNVFSC--SVNHEALHNHYTQKSLSLSP-----GLQDDECAEADGELDGLMT 645
DB 2416 SNVYGMASQSVNLSVHGPPYTVLPREGVHYVMKMDITLICTSSGEPSPSPRWT 2469

RESULT 66
NEO1_MOUSE STANDARD; PRT; 1493 AA.
AC P97758;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neogenin precursor.
GN NEO1 OR NGN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Brain;
RX MEDLINE=97407661; PubMed=9264410;
RA Keeling S.L., Gad J.M., Cooper H.M.;
RT "Mouse neogenin, a DCC-like molecule, has four splice variants and is

```

```

RT expressed widely in the adult mouse and during embryogenesis."
RL Oncogene 15:691-700(1997).
CC -1- FUNCTION: May be involved as a regulatory protein in the
CC transition of undifferentiated proliferating cells to their
CC differentiated state. May also function as a cell adhesion
CC molecule in a broad spectrum of embryonic and adult tissues.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoform=5;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=P97798-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P97798-2; Sequence=VSP_002594;
CC Name=3;
CC IsoId=P97798-3; Sequence=VSP_002595;
CC Note=Expression developmentally regulated;
CC Name=4;
CC IsoId=P97798-4; Sequence=VSP_002596;
CC Note=Expression developmentally regulated;
CC Name=5;
CC IsoId=P97798-5; Sequence=VSP_002597;
CC Note=Expression developmentally regulated;
CC -1- TISSUE SPECIFICITY: Widely expressed.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED UBICUOUSLY THROUGHOUT THE MID TO
CC LATE STAGES OF GESTATION AND IN ADULT TISSUES. STRONG EXPRESSION
CC IS OBSERVED IN THE VENTRAL REGION OF THE VENTRICULAR ZONE OF THE
CC E15.5 MOUSE NEURAL TUBE, AS WELL AS IN THE VENTRICULAR ZONES OF
CC THE MESENCEPHALON AND RHOMBENCEPHALON. ISOFORMS 3 AND 4 ARE
CC EXPRESSED AT HIGHER LEVEL COMPARED TO OTHER ISOFORMS BETWEEN E11.5
CC AND E16.5.
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily, DCC family.
CC -1- SIMILARITY: Contains 4 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 6 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y09535; CA07027.1; -.
CC HSSP; P02751; ITTF.
CC DR MED; MG1.1097159; NEO1.
CC DR InterPro; IPR008957; FN_III-like.
CC DR InterPro; IPR003961; FN_III.
CC DR InterPro; IPR003962; FNIII subd.
CC DR InterPro; IPR007110; Ig-like.
CC DR InterPro; IPR003598; Ig_C2.
CC Pfam; PF00041; fn3; 6.
CC DR Pfam; PF00041; Ig; 4.
CC DR PRINTS; PR00014; FNTYPEIII.
CC DR SMART; SM00408; IGC2; 6.
CC DR SMART; SM00408; IGC2; 4.
CC DR PROSITE; PS50835; IG_LIKE; 4.
CC Cell adhesion; Repeat; Signal; Transmembrane; Immunoglobulin domain;
CC Glycoprotein; Alternative splicing.
CC KW SIGNAL 1 36
CC FT CHAIN 37 1493
CC FT DOMAIN 37 1136
CC FT TRANSMEM 1137 1157
CC FT DOMAIN 1158 1493
CC FT DOMAIN 63 158
CC FT DOMAIN 163 249
CC FT DOMAIN 254 347
CC FT DOMAIN 352 437
CC FT DOMAIN 467 564
CC FT DOMAIN 567 660
CC FT DOMAIN 661 760
CC FT DOMAIN 766 860
CC FT DOMAIN 861 981
CC FIBRONECTIN TYPE-III 5.

```



```

FT DOMAIN 982 1083 FIBROECTIN TYPE-III 6.
FT DISULFID 1149 1153 POLY-VAL.
FT DISULFID 85 140 BY SIMILARITY.
FT DISULFID 184 232 BY SIMILARITY.
FT DISULFID 281 331 BY SIMILARITY.
FT DISULFID 373 421 BY SIMILARITY.
FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 337 337 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 746 746 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 940 940 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 442 461 Missing (in isoform 2).
FT VARSPLIC 863 878 /Prid=VSP 002594.
FT VARSPLIC 1086 1096 /Prid=VSP 002595.
FT VARSPLIC 1279 1331 /Prid=VSP 002596.
FT VARSPLIC 1279 1331 Missing (in isoform 5).
SQ SEQUENCE 1493 AA; 163159 MW; 441DE919D5E17C0E CRC64;

Query Match 4.8%; Score 165.5; DB 1; Length 1493;
Best Local Similarity 21.8%; Pred. No. 0.0088;
Matches 159; Conservative 79; Mismatches 282; Indels 211; Gaps 40;

10 LLLVLTALL--PAA-----TQGNKV-----VLGKKGDTVELT 40
25 LLLLLLFLLLGLRFPASGAATKSGPRROGASVTRFTFPLVEPVDLTLSVSGSVILN 84
41 CTSASOKKSIOFMKNSNOIKILGNOGSLTKGSKINDADSRSLMDGNFLLIKNL- 99
85 CSAYSESPNIEK-----KDGTF-----LNLSDDKRLPPGS--LFTSNV 126
100 ----KIEDSTYIC--EVED---QKEEYQLVFGI---TANSDFHLQGSILTLTLES 145
127 HSKHNKPDGEGFOCVATVDNLGITVSRKATLVAGLPRTSQEPEPSVYVGSALINCEV 186
146 PEGSSPSV---QCRSP-----RKNIOGKTLVSQLELQDSGTWTCTVLQN-----QKK 192
187 NADLVFVFWEQNRQPLLDRLDRIVKLPSSG-TLVISNATGDDGLYRCIVESGGPKFSDE 245
193 VEFKI-----DIVVLAFOKASSIVYKKEGEQVEFSPLAFYVEKLTGSGSELMWQARAS 247
246 AELKVLQDEBEIVDLVFLMRPSSMKVYTO---SAVLCPVSGLPAPYVRW----- 293
248 SKSWITFDLKNKEVSKVKTODPKLQMGKKLPLHLTLPOALPOY---AGSGNLT----- 298
294 -----MKNEVLDTESSGRVLVLAGGLEISDVTEDDAGTYFCIALNGNKTYEAQNE 345
299 LALEAKTGLHQEVLNVNRATOLQKNLTCEVWGPTSPKML-----SLKLENKEAKVSR 354
346 LTVQVPGFLKQPRANIYAHESMDIV--FCEVETGKETPTVKNYKNGDVIIPSDNFXIYKE 403
355 E--KPYWLVNPEAKMGQCLL-SDSQ-----VLLESNIKVLPTSTVEPEKSCDKTIT 404
404 HNLQVIGLVKSDGEGFOCIAENDVGAQAQAOIILIEHDV-AIPT-LPPTSILSATTDIL 461
405 CPSPAPELLGSPSVLPFPKPKDTLMISTRPEVTCVVVDVSHEDDEVFNMYVDOVEVH 464
462 ARAITGP-----LPSAPRD-----VASLIVSTR---ILKTRTPASDPH 498
465 NAKTKREBOYNSTYRVSVLTVLHODMLNGKEYKSKVNSKALPABIEKTIIS----- 516
499 GD-----NLITYSVFTKEGVDR-----RVNTSQPGEMQVTIQLMLPATVY 540
517 --KAKGPREPOVYTLPPSRDELTKQVSL-----TCLVKGRFPPSIANVEWS---NQ 565
541 IFKVMAQNRHSGSESSAPLRVE--TOPEVOLPGAPNIRAYATSPISITVTWETPLSGNE 599
566 PENNYKTTPVLDSDGSFLYSLKLTVDKSRMOGQNVFSCSVNHEALHNHYTKSLSL--- 622

```

```

Db 600 IQ-NRK-----LYWEKGTDE--QDIVVSSHSTYINCL-KKYTEYSPRVVAY 643
Qy 623 ---SPGLQIDE 630
Db 644 NKRGGVSTOD 654

RESULT 67
ID LAC3_MOUSE STANDARD; PRT; 104 AA.
AC P01845;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig lambda-3 chain C region.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83014953; PubMed=6812053;
RA Selsing E., Miller U., Wilson R., Storb U.;
RT "Evolution of mouse immunoglobulin lambda genes."
RL Proc. Natl. Acad. Sci. U.S.A. 79:4681-4685(1982).
RN [2]
RP SEQUENCE OF 1-18 (MYELOMA PROTEIN CBPC-49).
RA Breyer R.M., Sauer R.T., Eisen H.N.;
RT "The variable region of mouse lambda-3 chains."
RL ICN UCLA Symp. Mol. Cell. Biol. 20:105-110(1981).
RN [3]
RP SEQUENCE OF 10-104 (MYELOMA PROTEIN CBPC-49 AND MONOCLONAL ANTIBODY
RX MEDLINE=81223782; PubMed=6165998;
RA Azuna T., Steiner J.A., Eisen H.N.;
RT "Identification of a third type of lambda light chain in mouse
RL immunoglobulin."
RL Proc. Natl. Acad. Sci. U.S.A. 78:569-573(1981).
RN [4]
RP -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J00585; AAB59670.1; -
DR PIR; B93922; LIMS.
DR HSP; P01842; 2MCG.
DR InterPro; IPR007110; IG-1like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00407; IGC1; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region.
FT NON_TER 1 1
FT DOMAIN 1 99
FT DISULFID 27 85 IG-LIKE.
FT DISULFID 103 103 INTERCHAIN (WITH HEAVY CHAIN).
SQ SEQUENCE 104 AA; 11371 MW; 83CEBCDD4AA348EF1 CRC64;

Query Match 4.8%; Score 163; DB 1; Length 104;
Best Local Similarity 34.3%; Pred. No. 0.0038;
Matches 35; Conservative 16; Mismatches 47; Indels 4; Gaps 2;

521 QPRE-PQVYTLPPSRDELTKQVSLTCLVKGFPSDIAVEMESNQGPENNYKTTPVLD 579
1 QPKSYPTLTMPSPSEELQENKATLVCLISNFPSPGCVVANKANGTPTTGGVDTSNPTKE 60

```

```

Oy 580 DGSFPLYSLTYDKSRMOQGNVFCSCVMHEALHNHYTOKSL 621
Db 61 DNKYMASFHLHTSDQMSHNSFTCOVH---GDTYKSLIS 99

RESULT 68
ILL1_HUMAN STANDARD; PRT; 213 AA.
AC P15814;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Immunoglobulin lambda-like polypeptide 1 precursor (Immunoglobulin-
DE related 14.1 protein) (Immunoglobulin omega polypeptide) (Lambda 5)
DE (CD179b antigen).
GN IGLL1 OR IGL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89315835; PubMed=2501791;
RA Hollis G.F., Evans R.J., Stafford-Hollis J.M., Kozmeyer S.J.,
RA McKearn J.P.;
RT "Immunoglobulin lambda light-chain-related genes 14.1 and 16.1 are
RT expressed in pre-B cells and may encode the human immunoglobulin
RT omega 1 light-chain protein."
RL Proc. Natl. Acad. Sci. U.S.A. 86:5552-5556(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Lymphoid;
RA MEDLINE=91108327; PubMed=1703205;
RA Evans R.J., Hollis G.F.;
RT "Genomic structure of the human Ig lambda 1 gene suggests that it may
RT be expressed as an Ig lambda 14.1-like protein or as a canonical B
RT cell Ig lambda 1 light chain: implications for Ig lambda gene
RT evolution."
RL J. Exp. Med. 173:305-311(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX TISSUE=Testis;
RA MEDLINE=22388257; PubMed=12477932;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Schenker C.M., Schuler G.D.,
RA Hopkins R.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stagleon M.J., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunnarac P.H.,
RA Richards S., Wozney K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield A.S., Schin J.E., Jones S.J.M., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SUBUNIT: Associates non-covalently with VPREB1.
RN [5]
RP TISSUE SPECIFICITY: EXPRESSED ONLY IN PRE-B-CELLS AND A SPECIAL
RN B-CELL LINE (WHICH IS SURFACE IG NEGATIVE).
RN [6]
RP SIMILARITY: Belongs to the immunoglobulin superfamily.
RN [7]
RP DATABASE: NAME=PROT; NOTE=PROT 1:64-67(2000);
RN WWW="http://www.ncbi.nlm.nih.gov/prov/guide/696419174_g.htm".
RN [8]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL Outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announce/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; M27749; AAA36100.1; -.
DR EMBL; M34513; AAA36096.1; -.
DR EMBL; M34511; AAA36096.1; JOINED.
DR EMBL; M34512; AAA36096.1; JOINED.
DR EMBL; BC012293; AAH12293.1; -.
DR PIR; A33911; A33911.
DR HSSP; P01842; 7PAB.
DR Genew; HGNC:5870; IGLL1.
DR MIM; 146770; -.
DR GO; GO:0016020; C:membrane; NAS.
DR GO; GO:0003793; F:defense/immunity protein activity; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig cl.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00407; IGL1; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW Antigen; Signal; Immunoglobulin domain.
FT SIGNAL 1 37 POTENTIAL.
FT CHAIN 38 213 IMMUNOGLOBULIN LAMBDA-LIKE POLYPEPTIDE 1.
FT DOMAIN 97 108 J REGION (BY SIMILARITY TO LAMBDA
FT DOMAIN LIGHT-CHAIN).
FT DOMAIN 109 213 C REGION (BY SIMILARITY TO LAMBDA
FT SEQUENCE 213 AA; 22963 MW; 9133A742B943C79 CRC64;
SQ
Query Match 4.8%; Score 162.5; DB 1; Length 213;
Best Local Similarity 30.5%; Pred. No. 0.0011;
Matches 36; Conservative 21; Mismatches 44; Indels 17; Gaps 3;
Oy 521 QPR-EPQVYTLPPSRDELTKNOVSLTCTVKGFPYSDIAYVESNGQP-ENNYKTPPVLD 578
Db 109 QPKATPSVTLPPSSSELOANKATLVCMDFPIGILITVWKADGTITQGVETTTTSKO 168
Oy 579 SDGSFPLYSLTYDKSRMOQGNVFCSCVMHEALHNHYTOKSLSPGLQDDECAEQ 636
Db 169 SNKYMASFHLHTSDQMSHNSFTCOVH-----GSTYEKTAAPAE 211

RESULT 69
CD22_HUMAN STANDARD; PRT; 847 AA.
AC P20273; Q95699; Q95701; Q95702; Q95703; Q01665; Q92872; Q92873;
AC Q9UQA7; Q9UQA8; Q9UQA9; Q9UQA0; Q9Y2A6;
DT 01-FEB-1991 (Rel. 17, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE B-cell receptor CD22 precursor (Lew-14) (B-lymphocyte cell adhesion
DE molecule) (BL-CAM) (Siglec-2).
GN CD22.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM CD22-BETA), AND VARIANT HIS-639.
RN [2]
RP TISSUE=Tonsil;
RX MEDLINE=91086838; PubMed=1985119;
RA Wilson G.L., Fox C.H., Fauci A.S., Kehr J.H.;
RT "cDNA cloning of the B cell membrane protein CD22: a mediator of B-B
RT cell interactions."
RL J. Exp. Med. 173:137-146(1991).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS CD22-ALPHA AND CD22-BETA).
RX MEDLINE=93267103; PubMed=8496602;

```

RA Wilson G.L., Najfeld V., Kozlow E., Menninger J., Ward D.,
 RA Kehrl J.H.;
 RT "Genomic structure and chromosomal mapping of the human CD22 gene.";
 RL J. Immunol. 150:5013-5024(1993).
 RN
 RN (3)
 RN SEQUENCE FROM N.A. (ISOFORMS CD22-ALPHA AND CD22-BETA).
 RA Lamedin J.E., McCreedy P., Adamson A.W., Butkhart-Schultz K.,
 RA Garcia E., Kyle A., Ramirez M., Stllwagen S., Garne J.,
 RA Danganan L., Bruce R., Quan G., Montgomery M., Ow D.,
 RA Kobayashi A., Olsen A.O., Carraro A.V.;
 RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
 RN
 RN (4)
 RN SEQUENCE FROM N.A. (ISOFORM CD22-BETA).
 RX MEDLINE=90231465; PubMed=1691828;
 RA Stamenkovic I., Seed B.;
 RT "The B-cell antigen CD22 mediates monocyte and erythrocyte adhesion.";
 RL Nature 345:74-77(1990).
 RN
 RN (5)
 RN SEQUENCE OF 13-137; 139-239; 241-328 AND 418-502 FROM N.A., AND
 RP VARIANTS THR-34; GLU-152; LYS-203; GLY-664; CYS-669 AND ASP-745.
 RX MEDLINE=99180618; PubMed=10079291;
 RA Hacta Y., Itochuya N., Matsushita M., Shiota M., Hagiwara K.,
 RA Tokunaga K.;
 RL "Identification of the gene variations in human CD22.";
 RT Immunogenetics 49:280-286(1999).
 RN
 RN (6)
 RN STALIC ACID BINDING.
 RX MEDLINE=93216636; PubMed=8463235;
 RA Powell L.D., Sgroi D., Sjoberg E.R., Stamenkovic I., Varki A.;
 RT "Natural ligands of the B cell adhesion molecule CD22 beta carry
 RT N-linked oligosaccharides with alpha-2,6-linked stalic acids that are
 RT required for recognition.";
 RL J. Biol. Chem. 268:7019-7027(1993).
 RN
 RN (7)
 RN INTERACTION WITH PTPN6.
 RX MEDLINE=95343349; PubMed=7618087;
 RA Doody G.M., Jutement L.B., Delibrias C.C., Matheve R.J., Lin J.,
 RA Thomas M.L., Pearson D.T.;
 RT "A role in B cell activation for CD22 and the protein tyrosine
 RT phosphatase SHP.";
 RL Science 269:242-244(1995).
 RN
 RN (8)
 RN INTERACTION WITH LYN; SYK AND PIK3R1/PIK3R2.
 RX MEDLINE=96257803; PubMed=8647200;
 RA Tusciano J.M., Engel P., Tedder T.F., Agarwal A., Kehrl J.H.;
 RT "Involvement of p72syk kinase, p53/56lyn kinase and phosphatidy1
 RT inositol-3 kinase in signal transduction via the human B lymphocyte
 RT antigen CD22.";
 RL Eur. J. Immunol. 26:1246-1252(1996).
 RN
 RN (9)
 RN INTERACTION WITH PTPN6; SYK AND PLCG1.
 RX MEDLINE=96195207; PubMed=8627166;
 RA Law C.L., Sldorenko S.P., Chandran K.A., Zhao Z., Shen S.H.,
 RA Fischer E.H., Clark E.A.;
 RT "CD22 associates with protein tyrosine phosphatase 1C, Syk, and
 RT phospholipase C-gamma(1) upon B cell activation.";
 RL J. Exp. Med. 183:547-560(1996).
 RN
 RN (10)
 RN REVIEW.
 RX MEDLINE=97288746; PubMed=9143697;
 RA Tedder T.F., Tusciano J., Sato S., Kehrl J.H.;
 RT "CD22, a B lymphocyte-specific adhesion molecule that regulates
 RT antigen receptor signaling.";
 RL Annu. Rev. Immunol. 15:481-504(1997).
 CC
 CC -1- FUNCTION: Mediates B-cell B-cell interactions. May be involved in
 CC the localization of B-cells in lymphoid tissues. Binds sialylated
 CC glycoproteins; one of which is CD45. Preferentially binds to
 CC alpha2,6-linked stalic acid. The stalic acid recognition site can
 CC be masked by cis interactions with stalic acids on the same cell
 CC surface. Upon ligand induced tyrosine phosphorylation in the
 CC immune response seems to be involved in regulation of B cell
 CC antigen receptor signaling. Plays a role in positive regulation
 CC through interaction with Src family tyrosine kinases and may also

CC act as an inhibitory receptor by recruiting cytoplasmic
 CC phosphatases via their SH2 domains that block signal transduction
 CC through dephosphorylation of signaling molecules.
 CC
 CC -1- SUBUNIT: Predominantly monomer of isoform CD22-beta. Also found as
 CC heterodimer of isoform CD22-beta and a shorter isoform. Interacts
 CC with PTPN6/SHP-1, LYN, SYK, PIK3R1/PIK3R2 and PLCG1 upon
 CC phosphorylation. Interacts with GRB2, INPP5D and SHC1 upon
 CC phosphorylation (By similarity). May form a complex with
 CC INPP5D/SHP-1, GRB2 and SHC1.
 CC
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Comment-Additional isoforms seem to exist;
 CC Name=CD22-beta;
 CC IsoId=P20273-1; Sequence=Displayed;
 CC
 CC Name=CD22-alpha;
 CC IsoId=P20273-2; Sequence=VSP_002531;
 CC
 CC -1- TISSUE SPECIFICITY: B lymphocytes.
 CC
 CC -1- DOMAIN: Contains 4 copies of a cytoplasmic motif that is referred
 CC to as the immunoreceptor tyrosine-based inhibitor motif (ITIM).
 CC This motif is involved in downmodulation of cellular responses.
 CC The phosphorylated ITIM motif binds to the SH2 domain of
 CC PTPN6/SHP-1.
 CC
 CC -1- PTM: Phosphorylated both on threonine/serine and tyrosine
 CC residues.
 CC
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. SIGLEC
 CC (stalic acid binding Ig-like lectin) family.
 CC
 CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
 CC
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC
 CC -1- CAUTION: Ref.4 sequence differs from that shown due to a
 CC frameshift in position 806
 CC
 CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD22 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd22.htm".
 CC
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X59350; CAA42006.1; -;
 CC EMBL; U62631; AAB06448.1; -;
 CC EMBL; U62631; AAB06449.1; -;
 CC EMBL; X52785; CAA36988.1; ALT FRAME.
 CC EMBL; AB012996; BAA36565.1; -;
 CC EMBL; AB012997; BAA36566.1; -;
 CC EMBL; AB012998; BAA36567.1; -;
 CC EMBL; AB012999; BAA36568.1; -;
 CC EMBL; AB013000; BAA36569.1; -;
 CC EMBL; AB013002; BAA36571.1; -;
 CC EMBL; AB013003; BAA36572.1; -;
 CC EMBL; AB013004; BAA36573.1; -;
 CC EMBL; AB013005; BAA36575.1; -;
 CC EMBL; AB013006; BAA36577.1; -;
 CC EMBL; HGNC:1643; CD22.
 CC
 CC GeneW; HGNC:1643; CD22.
 CC
 CC MIM; 107266; -;
 CC InterPro; IPR007110; IG-like.
 CC InterPro; IPR003598; IG_c2.
 CC Pfam; PF00047; IG_6
 CC SMART; SM00408; IGC2_4
 CC PROSITE; PS50835; IG_LIKE; 6.
 CC Cell adhesion; Lectin; Antigen; Transmembrane; Signal; Glycoprotein;
 CC Immunoglobulin domain; Repeat; Phosphorylation; Alternative splicing;
 CC Polymorphism.
 CC
 CC SIGNAL 1 19
 CC CHAIN 20 847 POTENTIAL.
 CC DOMAIN 20 867 EXTRACELLULAR (POTENTIAL).
 CC TRANSMEM 688 706 POTENTIAL.
 CC DOMAIN 707 847 CYTOPLASMIC (POTENTIAL).
 CC DOMAIN 847 847 IG-LIKE V-TYPE.
 CC DOMAIN 143 235 IG-LIKE C2-TYPE 1.

FT	DOMAIN	242	326	IG-LIKE C2-TYPE 2.
FT	DOMAIN	331	416	IG-LIKE C2-TYPE 3.
FT	DOMAIN	419	500	IG-LIKE C2-TYPE 4.
FT	DOMAIN	505	582	IG-LIKE C2-TYPE 5.
FT	DOMAIN	593	676	IG-LIKE C2-TYPE 6.
FT	SITE	760	765	ITIM MOTIF 1.
FT	SITE	794	799	ITIM MOTIF 2.
FT	SITE	820	825	ITIM MOTIF 3.
FT	SITE	840	845	ITIM MOTIF 4.
FT	DISULFID	39	167	BY SIMILARITY.
FT	DISULFID	44	102	BY SIMILARITY.
FT	DISULFID	161	219	BY SIMILARITY.
FT	DISULFID	265	309	BY SIMILARITY.
FT	DISULFID	353	396	BY SIMILARITY.
FT	DISULFID	442	484	BY SIMILARITY.
FT	DISULFID	529	571	BY SIMILARITY.
FT	DISULFID	616	659	BY SIMILARITY.
FT	MOD_RES	762	762	PHOSPHORYLATION
FT				(INVOLVED IN BINDING TO SYK)
FT				(BY SIMILARITY).
FT	MOD_RES	807	807	PHOSPHORYLATION
FT				(INVOLVED IN BINDING TO GRB2)
FT				(BY SIMILARITY).
FT	MOD_RES	822	822	PHOSPHORYLATION
FT				(INVOLVED IN BINDING TO SYK)
FT				(BY SIMILARITY).
FT	MOD_RES	842	842	PHOSPHORYLATION
FT				(INVOLVED IN BINDING TO SYK, PLCG2 AND
FT				PIK3R1/PIK3R2) (BY SIMILARITY).
FT				N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	67	67	

Query Match	4.8%	Score 162.5	DB 1	Length 847
Beet Local Similarity	19.4%	Pred. No. 0.0065		
Matches 132	Conservative 103	Mismatches 229	Indels 217	Gaps 36

```

QY 11INKLKIEDSDTYICE-----VEDQKEEVOQLVFGLTANSDFTL 133
Db 11LVLEELIAPSDSKKWFEPHPETLLYANEGACWMIPTCYRALDGLDESFFILPHNEPYKNKNSK 70
QY 134 LOGOSILTLE--SPGSSPSVQCRSPKGNIOGGKTLVSQLEILOSDESTWTCYLOXK 191
Db 71 FDGRLRYESTDGDKVPSEQKRVQFGLGDKKNC---TLSIHVHLANDSQ--LGIRMES 123
QY 192 KVEFKIDIVL-----AFQKASSIVYKKEGQVE-----ESFPLAFTVEKLT 233
Db 124 KTEKMEKHILNWSERPPPHQLPPEI---QESQEVTLTGLNNSCYGPI----- 172
QY 234 GSGELMWOAEPRASSSKSWITPDLKNKEVSXRVOTDPKIQMGKULPLHILTLPQALPYAG 293
Db 173 ---QLOMILEGVPMRQAAVT---STSLTIKSVFRSELKFS-----FQMSH 212
QY 294 SGNLTLALEAKTGL--HQEVNLVYMAITQLO-----KNLTCEWGWGPRS 335
Db 213 HGKIVTCOLOPADGFLSNDTVQNLVWKHTPKLEIVTSPDAIVREGDSVYTMCEV--SSSN 271
QY 336 PKLM-----LSIKLENKEAKSKREKPVNLVLPBAGMOCLSD-----SQOYLLE 381
Db 272 PEYTTWSMLKDGSTLSKKON--TFTLNLRE---VTKDQSGKCCQYNSNDVGPGRSESEVFLQ 326
QY 382 -----SNIKVLP-----TW-----SPVPEKSGCDKTHT--- 404
Db 327 VOYAPBETVOQLHSPAYEGSVEFLCMSLNPFLPTNYTHWNGEMOGRTBEKXHIKPI 386
QY 405 -----CPCPAPBELLG---GPSVFL---FPKPKDOLMTSRTD---EYTCVVVDVSH 447
Db 387 LPMHAGTVSCVAENILGTQGRGPGALDVOYPPKATVYIQNPMIRISGDTYTLSCONNS 446
QY 448 EDPEV-KNNYVYDGVFNNAKTKPREEOQNSYTYRVVSVLTVLHQDLNGEKYKCVSKN- 505
Db 447 SNPSYTRJEW-----KPHGAMEEPS-----LGVLIQONVGMDN--TTIACAACNSM 490
QY 506 -ALPAPIKITSKAKGQPREPOVYTLPPSRDELTKNOVSLTCLVGVGFPSDLAVEMESNG 564

```

Db 491 CSMASPAVALNOYA---PRDVRVRKIKPLSEIHSNGSNVSLQCFDFSSPHKEVQFFMEKNG 547

QY 565 -----QENNNKTPPVLDSDGSFPLYKRLTYDKRMQGNFSSGVNHEALHNHTQKS 619

Db 548 RLKGKESQUNPDSISF--EDAGSY-----SCVNNNSIGQASKAMTLEVLVYAPRR 595

QY 620 L--SLSPGLQLDE-----TC 632

Db 596 LRVSMSPGDQVMEGKSAITLTC 616

RESULT	70
VCAM1_HUMAN	
ID	VCAM1_HUMAN
AC	P19320;
DT	01-NOV-1990 (Rel. 16, Created)
DT	01-NOV-1990 (Rel. 16, Last sequence update)
DT	10-OCT-2003 (Rel. 42, Last annotation update)
DE	Vascular cell adhesion protein 1 precursor (V-CAM 1) (CD106 antigen)
DE	(NCAM1-100).
GN	VCAM1 OR L1CAM.
OC	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX	NCBI_Taxid=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Umbilical vein;
RX	MEDLINE=91016951; PubMed=1699207;
RA	Polte T., Newman W., Gopal T.V.;
RT	"Full length vascular cell adhesion molecule 1 (VCAM-1).";
RL	Nucleic Acids Res. 18:5901-5901(1990).
RN	[2]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=90090619; PubMed=2688899;
RA	Osborn L., Hession C., Tizard R., Vassallo C., Luhnowsky S.,
RA	Chi-Rosso G., Lobb R.;
RT	"Direct expression cloning of vascular cell adhesion molecule 1, a
RT	cytokine-induced endothelial protein that binds to lymphocytes.";
RL	Cell 59:1203-1211(1989).
RN	[3]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=91352090; PubMed=1715583;
RA	Cybulsky M.I., Fries J.W.U., Williams A.J., Sultana P., Eddy R.,
RA	Byers M., Shows T., Gimbrone M.A. Jr., Collins T.;
RT	"Gene structure, chromosomal location, and basis for alternative mRNA
RT	splicing of the human VCAM1 gene.";
RL	Proc. Natl. Acad. Sci. U.S.A. 88:7859-7863(1991).
RN	[4]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=91201302; PubMed=1707873;
RA	Hession C., Tizard R., Vassallo C., Schiffer S.B., Goff D., Moy P.,
RA	Chi-Rosso G., Luhnowsky S., Lobb R., Osborn U.;
RT	"Cloning of an alternate form of vascular cell adhesion molecule-1
RT	(VCAM1).";
RL	J. Biol. Chem. 266:6682-6685(1991).
RN	[5]
RP	SEQUENCE FROM N.A., AND VARIANTS PHE-318; ALA-384; ALA-413 AND
RP	LEU-716.
RA	Rider M.J., Armel T.Z., Carrington D.P., Ozuna M., Kuldanek S.A.,
RA	Rajkumar N.R., Toth E.J., Yi Q., Nickerson D.A.;
RL	Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
RN	[6]
RP	SEQUENCE FROM N.A. (ISOPFORM LONG).
RC	TISSUE=Retinal pigment epithelium;
RX	MEDLINE=22388257; PubMed=12477932;
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA	Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heib F.,
RA	Dietcheno L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA	Brownstein W.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodegren E.J., Lu X., Gibbs R.A.,
 RA Fahney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.M., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smalhus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Maira M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [7]
 RN X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 25-226.
 RP MEDLINE:95147978; PubMed:7531291;
 RX Jones E.Y., Harlow K., Bottomley M.J., Robinson R.C., Driscoll P.C.,
 RA Edwards R.M., Clements J.M., Dudgeon T.J., Stuart D.I.,
 RT "Crystal structure of an integrin-binding fragment of vascular cell
 RT adhesion molecule-1 at 1.8-A resolution.";
 RL Nature 373:539-544(1995).
 [8]
 RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 25-220.
 RP MEDLINE:95296382; PubMed:753925;
 RX Wang J.-H., Pepinsky R.B., Stehle T., Liu J.-H., Karpusas M.,
 RA Browning B., Osborn L.,
 RT "The crystal structure of an N-terminal two-domain fragment of
 RT vascular cell adhesion molecule 1 (VCAM-1): a cyclic peptide based on
 RT the domain 1 C-D loop can inhibit VCAM-1-alpha 4 integrin
 RT interaction.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:5714-5718(1995).
 [9]
 RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 25-220.
 RP Wang J.-H., Stehle T., Pepinsky R.B., Liu J.-H., Karpusas M.,
 RA Osborn L.,
 RT "Structure of a functional fragment of VCAM-1 refined at 1.9-A
 RT resolution.";
 RL Acta Crystallogr. D 52:369-379(1996).
 CC -1- FUNCTION: IMPORTANT IN CELL-CELL RECOGNITION. APPEARS TO FUNCTION
 CC IN LEUKOCYTE-ENDOTHELIAL CELL ADHESION. INTERACTS WITH THE BETA-1
 CC INTEGRIN VIA4 ON LEUKOCYTES, AND MEDIATES BOTH ADHESION AND SIGNAL
 CC TRANSDUCTION. THE VCAM1/VLA4 INTERACTION MAY PLAY A
 CC PATHOPHYSIOLOGIC ROLE BOTH IN IMMUNE RESPONSES AND IN LEUKOCYTE
 CC EMIGRATION TO SITES OF INFLAMMATION.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=Long;
 CC IsoId=PI9320-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=PI9320-2; Sequence=VSP_002580.
 CC -1- TISSUE SPECIFICITY: Expressed on inflamed vascular endothelium, as
 CC well as on macrophage-like and dendritic cell types in both normal
 CC and inflamed tissue.
 CC -1- INDUCTION: By cytokines (e.g. IL-1, TNF-alpha).
 CC -1- PTM: Sialoglycoprotein.
 CC -1- DISEASE: May play an important role in the genesis of
 CC atherosclerosis and rheumatoid arthritis.
 CC -1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.
 CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD106 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd106.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X53051; CAA57218.1; -
 CC EMBL; M30257; AAA51917.1; ALT_TERM.
 CC DR

DR EMBL; M73255; AAA61270.1; -
 DR EMBL; M60335; AAA61269.1; -
 DR EMBL; AF536818; AAA965190.1; -
 DR EMBL; BC01276; AAH1276.1; -
 DR PIR; A41288; A41288.
 DR PIR; B41288; B41288.
 DR PDB; 1VCA; 15-SEP-95.
 DR PDB; 1VSC; 20-JUN-96.
 DR PDB; 1I09; 07-NOV-01.
 DR Genew; HGNC:12663; VCAM1.
 DR MIM; 192225; -
 DR InterPro; IPR003987; ICAM VCAM-1.
 DR InterPro; IPR007110; IG-1-like.
 DR InterPro; IPR003598; IG_C2
 DR InterPro; IPR003989; VCAM-1.
 DR Pfam; PF00047; Ig_6.
 DR PRINTS; PR01472; ICAMVCAM1.
 DR PRINTS; PR01474; VCAM1.
 DR SMART; SM00408; IGC2; 3.
 DR PROSITE; PS50835; IG_LIKE; 5.
 KW Immunoglobulin domain; Glycoprotein; Cell adhesion; Transmembrane;
 KW Repeat; Signal; Alternative splicing; Polymorphism; 3D-structure.
 FT SIGNAL 1 24
 FT CHAIN 25 739
 FT DOMAIN 25 698
 FT TRANSMEM 699 720
 FT DOMAIN 721 739
 FT DOMAIN 25 105
 FT DOMAIN 109 212
 FT DOMAIN 223 309
 FT DOMAIN 312 399
 FT DOMAIN 408 506
 FT DOMAIN 511 595
 FT DOMAIN 600 684
 FT DISULFID 47 95
 FT DISULFID 52 99
 FT DISULFID 137 195
 FT CARBOHYD 273 273
 FT CARBOHYD 365 365
 FT CARBOHYD 417 417
 FT CARBOHYD 463 463
 FT CARBOHYD 531 531
 FT CARBOHYD 561 561
 FT VARPEPIC 310 402
 FT VARIANT 318 318
 FT VARIANT 384 384
 FT VARIANT 413 413
 FT VARIANT 716 716
 FT STRAND 26 30
 FT STRAND 34 38
 FT TURN 39 40
 FT STRAND 43 50
 FT STRAND 56 61
 FT TURN 62 63
 FT STRAND 70 74
 FT TURN 75 76
 FT STRAND 77 82
 FT HELIX 87 89
 FT STRAND 92 99
 FT TURN 100 101
 FT STRAND 102 114
 FT STRAND 120 123
 Query Match 4.7%; Score 161.5; DB 1; Length 739;
 Best Local Similarity 21.5%; Pred. No. 0.0063;
 Matches 118; Conservative 91; Mismatches 217; Indels 123; Gaps 27;
 8 RHLLVLQDLALPLPAT--QGKKVVLGKGGDVELTCTASQKSKSIFPHKNSNQIKILNQ 65

```

Db      211 ROAVELGVYVISPKNVTVISVNPSTKLQEGGCVTWCSEGLPAPRIFW----- 258
Qy      66 GSFLLKSGKUNDRADSRSLMDGNFPLIKNKLIEPDRYICG-----VEDQKEVOLL 121
Db      259 -----SKKIDNGNLQHL---SGNATLTLLAMMDSGIVYCGVNLIGKNRKEVELI 307
Qy      122 V-----FGLTANSDTHLQ---GQSLTLTLTSPGSSPSVQCSPPGKNIQ-----GKT 168
Db      308 VQEKFTVEISPGRIAAQIGDSVWLTCESVWGCSFSPSMRTQIDISPLSGKVRSGTNT 367
Qy      169 LSVGLELQDSGTWTCTYVLQNKXVEFKIDIVLAFQKASSIVK---KEGEQVEFS--- 222
Db      368 LTLSSVSEFENEHSYLCYTCGKKLEKGIQVELYSFPDPRIEMSGLVNSSVTVSCKV 427
Qy      223 ---PFL-AFTVEKLTGSGELMWQARASSSKSWTF-----DLKNEVSVKRTYQDPKIQ- 273
Db      428 PSVYFLDRLEITLNGETIL-----ENIFFLEDTPDKSLKNSLEMTFIPTIED 476
Qy      274 MGKKL-----PLHLTLPOALPOYAGSGNLTALAEATGKLHOFNLVVRATOLQK---N 325
Db      477 TGKALVCAKHIIDMEFEPRQRS---TQTLVNVAP--RRTYLVSPSSILEGSSVN 531
Qy      326 LTCEWGPSPKMLSLKLENKAVSKREKPVVLN---PRAGMQL-LSDGQVLE 381
Db      532 MTCLSQGFPAPKILMSRQLPNEGLOPLSENATLTLISTMDESGVYLCEGINQGRSRKE 591
Qy      382 SNIKVLPFTWSTFVEKSCDKTHTCPCPAPELLGGPSVFLPDS---KPKDTLMSRT--- 435
Db      592 VELIT-----QVTKDKIKLT-----APFSEVSKGDVYIISCTCGN 627
Qy      436 -PEVTGVVDVSHEDPEVKFMYVDGV-EVHNAKTKPRB---EQYSTYRVSVLYTLAH 489
Db      628 VPE-TWILKKKAEFTGDTVLK-SIDGAYTIRKAQKDGAVYECESKKNVGSQRLSLTLDV 685
Qy      490 QDMNGKEY 498
Db      686 QGRNNKDY 694

RESULT 71
LAC2_MOUSE
ID LAC2_MOUSE STANDARD; PRT; 104 AA.
AC P01844;
DT 21-JUL-1986 (Rel. 01, Created)
RT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
OS Ig lambda-2 chain C region.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83014953; PubMed=6812053;
RA Seising E., Miller J., Wilson R., Storb U.;
RT "Evolution of mouse immunoglobulin lambda genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:4681-4685(1982).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=8274221; PubMed=6287422;
RA Wu G.E., Govindji N., Hozumi N., Murialdo H.;
RT "Nucleotide sequence of a chromosomal rearranged lambda 2
  immunoglobulin gene of mouse.";
RL Nucleic Acids Res. 10:3831-3843(1982).
RN [3]
RP SEQUENCE FROM N.A. (MOPC 315).
RX MEDLINE=82220143; PubMed=6283385;
RA Baltsmore D.;
RA Baltsmore D.;
RT "Somatic variants of murine immunoglobulin lambda light chains.";
RL Nature 298:380-382(1982).
RN [4]

```

```

RP SEQUENCE (MOPC 315).
RX MEDLINE=74048693; PubMed=4760498;
RA Dugan E.S., Bradehaw R.A., Simms E.S., Eisen H.N.;
RT "Amino acid sequence of the light chain of a mouse myeloma protein
  (MOPC-315).";
RL Biochemistry 12:5400-5416(1973).
RN [5]
RP SEQUENCE OF 66-104 (MOPC 315), AND REVISIONS.
RX MEDLINE=8123782; PubMed=6165998;
RA Azuma T., Steiner L.A., Eisen H.N.;
RT "Identification of a third type of lambda light chain in mouse
  immunoglobulins.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:569-573(1981).
CC -1-SIMILARITY: Contains 1 immunoglobulin-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.ebi.ch/announce/
  or send an email to license@ebi.ebi.ch).
CC
CC EMBL; J00595; AAA39151.1; -.
CC PIR; G93922; L2MS.
CC HSSP; P01842; 2MCG.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003597; Ig c1.
CC InterPro; IPR003006; Ig_MHC.
CC Pfam; PF00047; Ig; 1.
CC SMART; SM00407; Igcl; 1.
CC PROSITE; PS50835; IG_LIKE; 1.
CC PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region.
FT NON TER 1
FT DOMAIN 1 99 IG-LIKE.
FT DISULFID 27 85
FT DISULFID 103 103 INTERCHAIN (WITH HEAVY CHAIN).
SQ SEQUENCE 104 AA; 11254 MW; CE4B67B868862D3 CRC64;

Query Match 4.7%; Score 161; DB 1; Length 104;
Best Local Similarity 34.3%; Pred. No. 0.00051;
Matches 35; Conservative 15; Mismatches 48; Indels 4; Gaps 2;

Qy 521 QPRE-PQVYTLPPSRDELTKQVSLTCLVKGFPYSDIAVENESGQENNYKTPPYLDS 579
Db 1 QPKSTPLTYVFPPESEELKKNKATLVCLISNFSSGTVAMKANGTITITGVDSINPTKE 60
Qy 580 DGSFLLYSKLTVDKSRMOQGVSFSCVMEALNNHYQKSLIS 621
Db 61 GNKFMASSFHLTSTDQWRSHNSFCQVTHE---GDVTEKSLIS 99

RESULT 72
HB2D_PIG
ID HB2D_PIG STANDARD; PRT; 258 AA.
AC P15983;
DT 01-APR-1990 (Rel. 14, Created)
RT 01-APR-1990 (Rel. 14, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
OS SLA class II histocompatibility antigen, DQ haplotype D beta chain
  precursor.
OC Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
OX NCBI_TaxID=9623;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90361905; PubMed=2391424;
RA Gustafsson K., Leguenn C., Hirsch F., Germana S., Pracht K.,
RA Sachs D.H.;
RT "Class II genes of miniature swine. IV. Characterization and
  expression of two allelic class II DQB cDNA clones.";
RN [4]

```



```

DR Pfam; PF00047; Ig; 6.
DR PROSITE; PS50835; IG_LIKE; 6.
KW Immunoglobulin domain; Glycoprotein; Lipoprotein; GPI-anchor;
KW Membrane; Signal; Repeat; 3d-structure.
FT SIGNAL 1 34
FT CHAIN 35 685
FT PROPER 686 702 ADHESION MOLECULE 5.
FT DOMAIN 35 144 REMOVED IN MATURE FORM (POTENTIAL).
FT DOMAIN 146 237 IG-LIKE 1.
FT DOMAIN 238 322 IG-LIKE 2.
FT DOMAIN 324 415 IG-LIKE 3.
FT DOMAIN 416 498 IG-LIKE 4.
FT DOMAIN 502 593 IG-LIKE 5.
FT DOMAIN 594 677 IG-LIKE 6.
FT LIPID 685 685 GPI-anchor amidated alanine (Potential).
FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 115 115 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 152 152 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 182 182 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 204 204 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 256 256 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 274 274 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 292 292 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 360 360 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 480 480 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 508 508 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 529 529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 553 553 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 560 560 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 580 580 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 612 612 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 650 650 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 665 665 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 320 320 MISSING (IN REF. 4).
SQ SEQUENCE 702 AA; 76795 MW; 6299AE26CDDDB5C CAC64;

Query Match 4.7%; Score 161; DB 1; Length 702;
Best Local Similarity 17.7%; Pred. No. 0.0064;
Matches 119; Conservative 108; Mismatches 255; Indels 190; Gaps 32;

OY 20 PATQGNKVVVKGKGDVETLCTASQKSIQFHMKNISQIKILNQSGFLTKGSKLNDK 79
DB 147 PSISNSNKPVEDK-DAVAFTCEPTQATYLMVNNOSLPV----- 187
OY 80 ADSRRSLWDOGNFPLIKLKIEDSDTYICEVED-----QKEEYQL-LVFGILTANS- 129
DB 188 --SPRLQISNGNRRLTLFNTRYNDPTASYKCEQNQVPSARRSDSYLVNLYGPDAPTISPL 245
OY 130 DTHLQOSLTLTLSPSGSSPSVQCR-SPRGKNIQSGKTVSGLQELQDSGTTCTVQLQ 188
DB 246 NTSYRSNGNLMISCA--ASNPPQVYSFNVGTRQOSTQELFIRITVNNNGSYTCQAHN 303
OY 189 NQKKV-EFKIDIVVLAFOKASSIYKKGEQVRSFPLAFVE-KLTGSGELMQABRAS 246
DB 304 SDTGLNRTTITVTVYAEPPFITSNNNSNVEDEDAVALTCEPEIQNTYLLM----- 357
OY 247 SSKSMITFDLKNKEVSVKRVTDPRQLQWKKLPLHLPLQALPOYAGSGNLTALAEKATG 306
DB 358 -----VNNOSLPV-----SPRLQL-----SNDNRTLVLLSVTR 385
OY 307 KLAHEVNLVYKRAQLQKNTLCE-----VNGPSP-----KMLSLKL 344

```

```

DB 386 -----NDVSPYECGIGNELSDVSDPVLNLYGPDPTISPSYTYRPGVNLISLCHA 439
OY 345 ENK-----BAKYSKREKPPVVLN---PEAGMQCLSDSGQVLLSEINIKVLPMTST 392
DB 440 ASNPAPQSWLIDONIQOHQELFISNTEKNSGLYTCQANNSAGSRITVKTI-TVSA 498
OY 393 PVEBKSCDKTHTCPCPAPELLGQPSVLFPPPKKDTLMTSRPEVTCVVVDVSHEDPEV 452
DB 499 EL-----PKPSISNSNKR---FVEDKDAVAFTCEPEA-----QNT 530
OY 453 KFMVYVGVENVNAKTPREEQVNSTYRVSVLTVLHQDPLNGEYKCVSNKALPAPIE 512
DB 531 TYLMVNG---QSLPVSRLQISNGN-RTLLEFVTRND---AAAYVCGIQL----- 575
OY 513 KTSKAGQPREPQV-----YTLPPSRDELTFKQVSLTCLVKGFPSPDIAMWESNQ 565
DB 576 -SVANSRSDVTLVDLYGPPTPIISPDSLSYLSGANLNLSC-HSASNPSP-QYSWRINGI 632
OY 566 PENNYKTPPVLDSDGSFFLYSKLTYDKSRWQGNVFSGVMEBAL-HNHYTKSLSL-- 622
DB 633 PQCHTQV-----LFTAKITPNN-----GTVACFVSNLATGRNNSIVKSIITVSA 676
OY 623 ---SPGLQDET 631
DB 677 SGTSFGLSAGAT 688

RESULT 74
LAC_HUMAN STANDARD; PRT; 105 AA.
ID_P01842; P80423;
AC 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE IG_Lambda chain C regions.
GN IG_L1 AND IG_L2 AND IG_L3.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE (BENCE-JONES PROTEIN SH).
RX MEDLINE=70166723; PubMed=4909564;
RA Titeani K., Wikler M., Shinoda T., Putnam F.W.;
RT "The amino acid sequence of a lambda type Bence-Jones protein. 3. The
RT complete amino acid sequence and the location of the disulfide
RT bridges.";
RL J. Biol. Chem. 245:2171-2176(1970).
RN [2]
RP SEQUENCE (BENCE-JONES PROTEIN X), AND DISULFIDE BONDS.
RX MEDLINE=69088380; PubMed=4883841;
RA Mistein C., Clegg J.B., Jarvis J.M.;
RT "Immunoglobulin lambda-chains. The complete amino acid sequence of a
RT Bence-Jones protein.";
RL Biochem. J. 110:631-652(1968).
RN [3]
RP SEQUENCE (BENCE-JONES PROTEIN NIG-64).
RX MEDLINE=63186114; PubMed=6404900;
RA Kamezani F., Takayasu T., Suzuki S., Shinoda T., Okuyama T.;
RT "Comparative studies on the structure of the light chains of human
RT immunoglobulins. IV. Assignment of a subgroup.";
RL J. Biochem. 93:421-429(1983).
RN [4]
RP SEQUENCE (BENCE-JONES PROTEIN KERN).
RX MEDLINE=71150336; PubMed=5549568;
RA Postelngl H., Hess M., Hilschmann N.;
RT "Structural rule of antibodies. Primary structure of a monoclonal
RT immunoglobulin-L-chain of the lambda type, subgroup IV (Bence-Jones-
RT protein Kern). V. The complete amino acid sequence and its genetic
RT interpretation.";
RL Hoppe-Seyler's Z. Physiol. Chem. 352:247-266(1971).
RN [5]

```


RP PARTIAL SEQUENCE (MYELOMA PROTEIN NEMM).
RX MEDLINE=74109253; PubMed=4814727;
RA Chen B.L., Poljak R.J.;
RT "Amino acid sequence of the (lambda) light chain of a human myeloma
RT immunoglobulin (IgG New).";
RL Biochemistry 13:1295-1302(1974).
RN [6]
RP SEQUENCE (DOT).
RX MEDLINE=95255298; PubMed=7737190;
RA Scoppin M., Bellotti V., Negri A., Merlini G., Garver F., Ferri G.;
RT "Characterization of the two unique human anti-Flavlin monoclonal
RT immunoglobulins.";
RL Eur. J. Biochem. 228:886-893(1995).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF NEMM.
RX MEDLINE=75046825; PubMed=4215080;
RA Poljak R.J., Amzel L.M., Avey H.P., Chen B.L., Phizackerley R.P.,
RA Saul F.;
RT "The three-dimensional structure of the Fab' fragment of a human
RT myeloma immunoglobulin at 2.0-A resolution";
RL Proc. Natl. Acad. Sci. U.S.A. 71:3440-3444(1974).
RN [8]
RP PARTIAL SEQUENCE (BENCE-JONES PROTEIN MCG).
RX MEDLINE=75013804; PubMed=4415202;
RA Felt J.W., Deutesch H.F.;
RT "Primary structure of the M_g lambda chain.";
RL Biochemistry 13:4102-4114(1974).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF MCG.
RA Edmundson A.B., Ely K.R., Abola E.E., Schiffer M.,
RA Panagiotopoulos N.;
RT "Rotational allotetramerism and divergent evolution of domains in
RT immunoglobulin light chains";
RL Biochemistry 14:3953-3961(1975).
RN [10]
RP X-RAY CRYSTALLOGRAPHY OF MCG.
RX MEDLINE=90133913; PubMed=2515285;
RA Ely K.R., Herron J.N., Harker M., Edmundson A.B.;
RT "Three-dimensional structure of a light chain dimer crystallized in
RT water. Conformational flexibility of a molecule in two crystal
RT forms.";
RL J. Mol. Biol. 210:601-615(1989).
RN [11]
RP SEQUENCE FROM N.A.
RX MEDLINE=82080680; PubMed=6273747;
RA Hieker P.A., Hollis G.F., Kormeyer S.J., Waldmann T.A., Leder P.;
RT "Clustered arrangement of immunoglobulin lambda constant region genes
RT in man.";
RL Nature 294:536-540(1981).
RN [12]
RP MISCELANEOUS: The sequence shown is the Kern-/Oz-/M_g-chain
CC found in proteins SH, X, and NIG-64. The Kern protein has the
CC Kern+ marker, the NEMM protein has the Oz+ marker, the M_g protein
CC has the Kern+ marker, and the M_g+ marker.
CC MISCELANEOUS: Six random lambda-type genes were identified and
CC the 3 most 5' were sequenced. These correspond to the M_g sequence
CC (lambda-1), the Kern-/Oz- sequence (lambda-2) and the Kern-/Oz+
CC sequence (lambda-3).
CC [13]
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sb-sib.ch).
CC -----
CC EMBL: J00253; AAA59107.1; -
CC EMBL: L28562; AAB36581.1; AAT_INIT.
CC EMBL: X51754; CAB38569.1; AAT_INIT.
CC EMBL: X51755; CAB36049.1; -
CC EMBL: X51755; CAA36051.1; -
CC PIR: A92057; L2HU.

DR PDB: 2MCG; 15-JUL-92.
DR PDB: 7FAB; 31-JAN-94.
DR PDB: 1AOK; 04-FEB-98.
DR PDB: 1LIL; 15-MAY-97.
DR Genew; HGNC:5855; IGLC1.
DR Genew; HGNC:5856; IGLC2.
DR Genew; HGNC:5857; IGLC3.
DR MIM: 147220; -.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006935; P:Immune response; NAS.
DR InterPro; IPR007110; IG-Like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig_c1; 1.
DR SMART; SM00407; IGL1; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; Bence-Jones protein;
KW 3D-structure.
FT NON_TER 1 1
FT DOMAIN 6 100 IG-LIKE.
FT DISULFD 27 86 INTERCHAIN (WITH HEAVY CHAIN).
FT DISULFD 104 104 A-> N (IN MCG+ MARKER).
FT VARIANT 5 5 /FTid=VAR_003898.
FT VARIANT 7 7 S-> T (IN MCG+ MARKER).
FT VARIANT 45 45 /FTid=VAR_003899.
FT VARIANT 45 45 S-> G (IN KERN+ MARKER).
FT VARIANT 56 56 /FTid=VAR_003900.
FT VARIANT 56 56 T-> K (IN MCG+ MARKER).
FT VARIANT 82 82 /FTid=VAR_003901.
FT VARIANT 82 82 R-> K (IN OZ+ MARKER).
FT STRAND 8 11 /FTid=VAR_003902.
FT HELIX 15 19
FT TURN 20 21
FT STRAND 24 32
FT STRAND 38 43
FT TURN 44 45
FT STRAND 46 48
FT TURN 52 59
FT STRAND 61 62
FT TURN 65 72
FT STRAND 75 80
FT HELIX 84 89
FT STRAND 84 89
FT STRAND 94 99
SQ SEQUENCE 105 AA; 11236 MW; DCD9C7C201C13CC2 CRC64;
Query Match 4.7%; Score 159; DB 1; Length 105;
Best Local Similarity 34.1%; Pred. No. 0.0007;
Matches 31; Conservative 18; Mismatches 40; Indels 2; Gaps 2;
OY 521 QPR-EPQVYTPPSRDELTKQVSLTCLVKGFPSDIAWVESNGQP-ENNYKTPPVLYD 578
DB 1 QPKAPSTLTPPSSSEELQANKATIVCLISDFPEAPVAVKAKAASSPVKAGVETTPSKQ 60
OY 579 SDGSFFLYSKLTVDKSRMOQGNVSCSYMHE 609
DB 61 SNKTYAASSYSLTPPEQWKSHRSYSCQVTHE 91
RESULT 75
LAC_PIG STANDARD; PRT; 105 AA.
AC P01846;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE IG lambda chain C region.
OS Sue scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxId=9823;

```

RN [1]
RP MEDLINE=78000254; PubMed=409425;
RA Novotny J., Franek F., Margolies M.N., Haber E.;
RT "Amino acid sequence of normal (microheterogeneous) porcine
RL Immunoglobulin lambda chains."
RT Biochemistry 16:3765-3772(1977).
CC -1- MISCELLANEOUS: This chain was obtained from a mixture of normal
CC immunoglobulins.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR: A02129; LIPG.
DR HSSP: P01842; 2MCG.
DR InterPro: IPR007110; IG-1-like.
DR InterPro: IPR003597; IG_C1.
DR InterPro: IPR003006; IG_MHC.
DR Pfam: PF00047; Ig_1.
DR SMART: SM00407; IGC1; 1.
DR PROSITE: PS50835; IG_LIKE; 1.
DR PROSITE: PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region.
FT DOMAIN 1 100 IG-LIKE.
FT DISULFID 27 86
FT DISULFID 104 104 INTERCHAIN (WITH HEAVY CHAIN).
SQ SEQUENCE 105 AA; 11003 MW; 3817AAEBD747C396 CRC64;

Query Match 4.6%; Score 156; DB 1; Length 105;
Best Local Similarity 35.2%; Pred. No. 0.0011;
Matches 32; Conservative 15; Mismatches 42; Indels 2; Gaps 2;

OY 521 QPR-EPQYVTLPPSRDELTKQVSIITCLVKGFPYSDIAVENESGQP-ENNYKTPPYLD 578
DB 1 QPKAAPVNLPPSSSEELGTNKAIVLCISFPYPAVTVWAKGTTVTCGVETTKSKQ 60
OY 579 SDGSFLYXKLTVDKSRWQGNVPSCSYME 609
DB 61 SNKTAASSYIALASDWMKSSSGFTQVTHE 91

RESULT 76
KMLS_CHICK STANDARD; PRT; 1906 AA.
AC P11799; P19038;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Myosin light chain kinase, smooth muscle and non-muscle isozymes
DE (EC 2.7.1.117) (MCK) [Contains: Telokin].
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM MCK-210).
RX MEDLINE=96033976; PubMed=7589469;
RA Wateranova D.V., Collings M., Lukae T.J., van Eldik L.J., Birukov K.G.,
RA Stepanova O.V., Shirinsky V.P.;
RT "Multiple gene products are produced from a novel protein kinase
RT transactin region."
RT FEBS Lett. 373:217-220(1995).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM MCK-108).
RX MEDLINE=90192792; PubMed=2315320;
RA Olson N.J., Pearson R.B., Needleman D.S., Hurwitz M.J., Kemp B.E.,
RA Means A.R.;
RT "Regulatory and structural motifs of chicken gizzard myosin light
RT chain kinase."
RT Proc. Natl. Acad. Sci. U.S.A. 87:2284-2288(1990).
RN [3]
RP SEQUENCE OF 649-1906 FROM N.A., AND PARTIAL SEQUENCE.
RX TISSUE=Fibroblastic; PubMed=2202734;
MEDLINE=90361738;

```

```

RA Shoemaker M.O., Lau W., Shattuck R.L., Kwiatkowski A.P.,
RA Matrisian P.E., Guerra-Santos L., Wilson E., Lukas T.J.,
RA van Eldik L.J., Waterson D.M.;
RT "Use of DNA sequence and mutant analyses and antisense
RT oligodeoxynucleotides to examine the molecular basis of nonmuscle
RT myosin light chain kinase autoinhibition, calmodulin recognition, and
RT activity."
RL J. Cell Biol. 111:1107-1125(1990).
RN [4]
RP SEQUENCE OF 1259-1906 FROM N.A.
RC TISSUE=Gizzard;
RX MEDLINE=87157587; PubMed=3030394;
RA Guerriero V. Jr., Russo W.A., Olson N.J., Putkey J.A., Means A.R.;
RT "Domain organization of chicken gizzard myosin light chain kinase
RT deduced from a cloned cDNA."
RL Biochemistry 25:8372-8381(1986).
RN [5]
RP SEQUENCE OF 1750-1906 FROM N.A. (ISOFORM TELOKIN).
RC TISSUE=Gizzard;
RX MEDLINE=93073972; PubMed=1444462;
RA Yoshikai S., Ikebe M.;
RT "Molecular cloning of the chicken gizzard telokin gene and cDNA."
RL Arch. Biochem. Biophys. 299:242-247(1992).
RN [6]
RP SEQUENCE OF 1750-1906 FROM N.A. (ISOFORM TELOKIN).
RX MEDLINE=92236611; PubMed=1373815;
RA Collinge M., Matrisian P.E., Zimmer W.E., Shattuck R.L., Lukas T.J.,
RA van Eldik L.J., Waterson D.M.;
RT "Structure and expression of a calcium-binding protein gene contained
RT within a calmodulin-regulated protein kinase gene."
RL Mol. Cell. Biol. 12:2359-2371(1992).
CC -1- FUNCTION: Phosphorylates a specific serine in the N-terminus of a
CC myosin light chain, which leads to the formation calmodulin-MCK
CC signal transduction complexes which allow selective transduction
CC of calcium signals.
CC -1- CATALYTIC ACTIVITY: ATP + [myosin light-chain] = ADP + [myosin
CC light-chain] phosphate.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative Initiation;
CC Comment=At least 3 isoforms, MCK-210/Non-muscle,
CC MCK-108/Smooth-muscle and Telokin, are produced by alternative
CC initiation;
CC -1- TISSUE SPECIFICITY: ISOFORM TELOKIN IS EXPRESSED IN GIZZARD,
CC HEART, LUNG, INTESTINE, AND SKELETAL MUSCLE ALTHOUGH THE LEVELS OF
CC THE EXPRESSION IN THE LATTER WERE MUCH LESS THAN THAT IN THE
CC GIZZARD.
CC -1- DOMAIN: TELOKIN BINDS CALMODULIN.
CC -1- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.
CC -1- SIMILARITY: Contains 9 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 fibronectin type III domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X52876; CA37056.1; -
DR EMBL: X52876; CA37057.1; -
DR EMBL: X52876; CA37058.1; -
DR EMBL: M31048; AAA49069.1; -
DR EMBL: M14953; AAA69964.1; -
DR EMBL: M96655; AAA48083.1; -
DR EMBL: M88283; AAA48647.1; -
DR EMBL: M88284; AAB53768.1; -
DR PIR: S68235; S68235.
DR PDB: 1CDL; 31-AUG-94.
DR PDB: 1VRK; 27-APR-99.
DR InterPro: IPR008957; FN_III-like.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR007110; IG-1-like.

```

Query Match	Best Local Similarity	Score 155; DB 1; Length 1906;
Matches 149; Conservative 90; Mismatches 256; Indels 212; Gaps 36		
InterPro: IPR003598; Ig_c2.		
InterPro: IPR000719; Prot_kinase.		
InterPro: IPR008271; Ser_thr_pkin_AS.		
InterPro: IPR002290; Ser_thr_pkinase.		
Pfam: PF00041; fn3; 1.		
Pfam: PF00047; Ig; 9.		
Pfam: PF00069; pkinase; 1.		
ProDom: PP000001; Prot_kinase; 1.		
SMART; SM00060; FN3; 1.		
SMART; SM00408; IGC2; 8.		
SMART; SM00220; S_TKc; 1.		
PROSITE; PS50835; IG_LIKE; 9.		
PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.		
PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.		
PROSITE; PS00108; PROTEIN_KINASE_ST; 1.		
Transferrase; Serine/threonine-protein kinase; Calmodulin-binding; ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat; Alternative initiation; 3D-structure.		
CHAIN	1	1906
FT		
FT	CHAIN	935 1906
FT		
FT	CHAIN	1750 1906
FT		
FT	INIT_MET	935 935
FT	INIT_MET	1750 1750
FT	DOMAIN	28 117
FT	DOMAIN	156 244
FT	DOMAIN	429 517
FT	DOMAIN	521 613
FT	DOMAIN	637 725
FT	DOMAIN	735 830
FT	DOMAIN	1084 1172
FT	DOMAIN	1225 1313
FT	DOMAIN	1330 1400
FT	DOMAIN	1453 1708
FT	DOMAIN	1794 1885
FT	NP_BIND	1459 1467
FT	BINDING	1482 1482
FT	ACT_SITE	1574 1574
FT	DOMAIN	1716 1728
FT		
FT	DOMAIN	1730 1749
FT		
FT	DOMAIN	1317 1364
FT	DOMAIN	1385 1402
FT	DOMAIN	660 1833
FT	REPEAT	660 676
FT	REPEAT	758 774
FT	REPEAT	1107 1123
FT	REPEAT	1817 1833
FT	DOMAIN	653 1866
FT	REPEAT	693 708
FT	REPEAT	791 807
FT	REPEAT	1140 1156
FT	REPEAT	1281 1297
FT	REPEAT	1851 1866
FT	DOMAIN	970 1226
FT	REPEAT	970 987
FT	REPEAT	999 1016
FT	REPEAT	1061 1078
FT	REPEAT	1209 1226
FT	DOMAIN	1700 1763
FT	DOMAIN	1896 1906
FT	MOD_RES	1748 1748
FT	MOD_RES	1762 1762
FT	CONFLICT	1439 1439
FT	SEQUENCE	1906 AA; 210445 MW; AD7D8AB69EE3363 CRC64;

QY	89	QGNPRLIKNIKIEDSDTICEVED-----QKEEYVLVFGITASDTHLIGQSGILTLE	144
Db	81	RGIFSLVITKGVGDSGSKYTCEAANDGVRQVYVELIVEG--NS-----LKYSLPSSAK	133
QY	145	SPPG--SSPEVQCR-----SPRGNIQGGKT	168
Db	134	TPEGRLSVPEVHRPSINGESPRAKTRPNRVVAREGGTGAFSCKITGRPOQPTWTKGD	193
QY	169	-----LSVSOLELQSDSGTWITCTVLONOKKVEFKIDIVY-----	201
Db	194	IHLQNERFMNFEKGTGIQYLEIQAIVQVLADAGIYCTCTVNSAGKASVAELTVQSPDKTDT	253
QY	202	-----LA-----FOKASIIYKKKEGEQVESFPFLATVEKLTCSSGLMW	244
Db	254	HAQPLCMPKPPKTLTAKLAINSDFFQKATNSNGIAELKSTSEL-MVEFKRLSKKKEFTY	312
QY	241	QAEASASSKSWITFDLQKNKESVSKVNTODPKLOMGKKLPLHLTLPOALPYQVAGSNLTLLA	300
Db	313	TSEKADKGQKQNGEANNVPLDESGTGKQPVQLQKTSITLQAVKQKQPEKAEPOTFI	372
QY	301	LEAKTGKTHQEVNLVVMARATQLOKMLTCEWNGPTSPKIMLSKLEN-----KEAKYSKRE	355
Db	373	RQAEHDRK-RTVQPLMTTTTQGENPSLT-----QQVSPR---SRETFENAGYRKSKVKEKRE	423
QY	356	KPVWVLNPEAKGMQCLSDSGQ-VLLESNIKLP---TW---STPEVPSKSC-----DK	401
Db	424	PLGIPPEPESRPOSLEABEGQEIFKFSKVSQKPEPDVEMKEGVPFKTEBGIQIYEEDG	482
QY	402	THHC---PPCPABELLIGSPSVLPFPKPKDPT---LMSIRPEVT-----CVVV	443
Db	483	THCLMLKALKLQDS--GSYSCAFAFPDGGTSTISMLTLVREKPEVEVAPCRSSVLKGTIV-	538
QY	444	DVSHEDPEVKFMNYVDGVVHNANKRPREE--QY-NSTYRV-VSVLTVLHODMWLNGKE--	497
Db	540	---SEGDGDFLQCYGVGVPEPITWMLNEQPIQVHNSTFEAGVAKLTV--QDALPEDDGI	594
QY	498	YKCKVSNKA-----LPALEKTIISAKGQPREPOVYTLTP-----SRDELTKNOV	542
Db	595	YTCLAEANNGRASCSAQVTVKKEKSKSAEAGTQAKLNKTKFAPFLKGLTDLTKMDSQV	654
QY	543	SLTCLVKGFPFSDIANEMESNOPE---NNYKTPPLVD-----SDGSFFLYSLKLTIV	591
Db	655	IMV-----VEVSANPCPEIIMLHNKEIQLETEDHFEKKGNSLYLQEVFPE	702
QY	592	DKSRMOQGNVFCSVNHEALHNHYQKSLSPGLQDDETCAEADG	638
Db	703	DTGK-----YTCEANNE-LGERTQYATL-----TYGEFQDG	732

RESULT 77

DTC_HUMAN

ID	DTC_HUMAN	STANDARD:	PRT:	383 AA.
AC	P01880;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Ig delta chain C region.			
GN	IGHD.			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_Taxid=9606;			
NP	[1]			
RP	SEQUENCE OF 1-162 (MYELOMA PROTEIN WAH).			
RX	MEDLINE=82082419; PubMed=6947220;			
RA	Putnam F.W., Takahashi N., Tetaert D., Debuire B., Lin L.-C.;			
RT	"Amino acid sequence of the first constant region domain and the			
RT	hinge region of the delta heavy chain of human IgG."			
RL	Proc. Natl. Acad. Sci. U.S.A. 78:6168-6172(1981).			
RN	[2]			
RP	SEQUENCE OF 158-383 (MYELOMA PROTEIN WAH).			
RX	MEDLINE=81237368; PubMed=6787589;			
RA	Lin L.-C., Putnam F.W.;			

```

RT "Primary structure of the Fc region of human immunoglobulin D:
RT implications for evolutionary origin and biological function.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:504-508(1981).
RN [3]
RP SEQUENCE OF 103-137 (MYELOMA PROTEIN NIG-65), AND CARBOHYDRATES.
RX MEDLINE=82231401; PubMed=7092891;
RA Takayasu T., Suzuki S., Kametani F., Takahashi N., Shinoda T.,
RA Okuyama T., Munekata E.;
RT "Amino acid sequence of galactosamine-containing glycopeptides in the
RT hinge region of a human immunoglobulin D.";
RL Biochem. Biophys. Res. Commun. 105:1066-1071(1982).
RN [4]
RP SEQUENCE OF 158-383 (MYELOMA PROTEIN NIG-65).
RX MEDLINE=81199406; PubMed=6785754;
RA Shinoda T., Takahashi N., Takayasu T., Okuyama T., Shimizu A.;
RT "Complete amino acid sequence of the Fc region of a human delta
RT chain.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:785-789(1981).
RN [5]
RP CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE=8113632; PubMed=7008791;
RA Takayasu T., Takahashi N., Shinoda T.;
RT "Amino acid sequence and location of the three glycopeptides in the
RT Fc region of human immunoglobulin D.";
RL Biochem. Biophys. Res. Commun. 97:635-641(1980).
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
DR PIR, A02175; DHU.
DR Genev. HGNC:5480; IGHD.
DR MIM, 147170;
DR GO, GO:0003823; F:antigen binding; NNS.
DR GO, GO:0006955; P:immune response; NNS.
DR InterPro, IPR007110; IG-1like.
DR InterPro, IPR003597; IG-cl.
DR InterPro, IPR003006; IG_MHC.
DR Pfam, PF00047; Ig_3.
DR SMART, SM00407; IGcl; 3.
DR PROSITE, PS50835; IG_LIKE; 3.
DR PROSITE, PS00290; IG_MHC; 2.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
FT NON TER 1 1
FT DOMAIN 6 98
FT 175 263 IG-LIKE 1.
FT DOMAIN 267 373 IG-LIKE 2.
FT DISULFID 15 15 IG-LIKE 3.
FT DISULFID 28 84 INTERCHAIN (WITH LIGHT CHAIN).
FT CARBOHYD 109 109 INTERCHAIN (WITH HEAVY CHAIN).
FT CARBOHYD 110 110 O-LINKED.
FT CARBOHYD 113 113 O-LINKED.
FT CARBOHYD 126 126 O-LINKED.
FT CARBOHYD 127 127 O-LINKED.
FT CARBOHYD 131 131 O-LINKED.
FT CARBOHYD 132 132 O-LINKED (POTENTIAL).
FT CARBOHYD 225 225 N-LINKED (GLCNAC. . .).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .).
FT CARBOHYD 367 367 N-LINKED (GLCNAC. . .).
SQ SEQUENCE 383 AA; 42125 MW; DAIDET3519CT6CID CRC64;
Query Match 4.5%; Score 154; DB 1; Length 383;
Best Local Similarity 23.2%; Pred. No. 0.008;
Matches 63; Conservative 34; Mismatches 92; Indels 82; Gaps 11;
Qy 407 PCPABELGSPVFLP-----PKKDTLMTSRPEVCVVVDVSHEDPEVKFMVYDGV 461
Db 2 PTKAPD-----VFPIISGRNRPD-----NSVVLACLI-TGYHFTSVTVYIM-GT 46
Qy 462 EVHNAKTRREQVNSTYRVSVLTFLHQMVLNGKCYKCVSN----- 504
Db 47 OSQPRTEPEIQRSRDSYVMTSQSLTPLOQWRQG-EYICVQVHTASKSKKEIFRWPSPK 105
Qy 505 -----KALPAPL-----EKTISSAKGQPRPQ----- 526
Db 106 AQAASVPTAQAEGSLAKATTAPATTNTGTGSGEKKKKKKERKEBERETKTDECPSH 165

```

```

Qy 527 -----VYTLPPSRDEL-TKNQVSLTCLVKGFPSDIAVWESNGQ-PENNYKTPPVLDSD 580
Db 166 QPLGVILTPAVQDLMDKATKFTCFVVGSDLDMDAHLTWAVAGVNPFGVBEGLLEHNSN 225
Qy 581 GSFFLYSKLTVDKSRWQGVFSCVMEAL 611
Db 226 GSQSQHSRLTPRSLMAGTSVCTLNLHPSL 256
RESULT 78
LACS_MUUSP STANDARD; PRT; 105 AA.
ID LACS_MUUSP
AC P20765;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig lambda-5 chain C region.
OS Mus spretus (western wild mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10096;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88196070; PubMed=3129289;
RA Mani F., Cazenave P.A., Kindt T.J.;
RT "Conservation of the immunoglobulin C lambda 5 gene in the Mus gene.";
RL EMBO J. 7:117-122(1988).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M35582; AAA39152.1; -.
DR HSSP, P01842; 2MCG.
DR InterPro, IPR007110; Ig-1like.
DR InterPro, IPR003597; IG-cl.
DR InterPro, IPR003006; IG_MHC.
DR Pfam, PF00047; Ig_1.
DR SMART, SM00407; IGcl; 1.
DR PROSITE, PS50835; IG_LIKE; 1.
DR PROSITE, PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region.
FT NON TER 1 1
FT DOMAIN 6 100
FT DISULFID 27 86 IG-LIKE.
FT DISULFID 104 104 INTERCHAIN (WITH HEAVY CHAIN).
SQ SEQUENCE 105 AA; 11674 MW; AAB417DF68471A17 CRC64;
Query Match 4.5%; Score 153.5; DB 1; Length 105;
Best Local Similarity 34.0%; Pred. No. 0.0016;
Matches 35; Conservative 20; Mismatches 43; Indels 5; Gaps 3;
Qy 521 QPR-EPOVYTLPPSRDELTKNQVSLTCLVKGFPSDIAVWESNGQ-PENNYKTPPVLD 578
Db 1 QPKSDPLVTLFPLSLKYLQANKYTLVCLVSEFPGTLVVMKVDGVPTGQVETTPSKQ 60
Qy 579 SDGSFFLYSKLTVDKSRWQGVFSCVMEALHNHYTQSL 621
Db 61 TNKRYWSYTLTLISDQMPHSRYSKRVTH---GNTVERSVS 100
RESULT 79
LACS_RABIT STANDARD; PRT; 105 AA.
ID LACS_RABIT
AC P01847;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)

```

```

DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE IG_Lambda chain C region.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
CC NCBI_TaxID=9986;
XX [1]
RN SEQUENCE.
RX MEDLINE=82091105; PubMed=6797414;
RA Garcia I., Jacou J.-C.;
RT "The primary structure of the constant region of Basilea-rabbit
RL Immunoglobulin lambda-chains."
RT Biochem. J. 197;177-183(1981).
CC -1- MISCELLANEOUS: This lambda chain expresses the c7 allotypic
CC specificity.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A02130; L7RB.
DR HSSP; P01842; 7FAB.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig1.
DR SMART; SM00407; IgC1; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; FALSE NEG.
KW Immunoglobulin domain; Immunoglobulin C region.
KW NON_TER
FT DOMAIN 1 100 IG-LIKE.
FT DISULFID 27 86
FT DISULFID 104 104 INTERCHAIN (WITH HEAVY CHAIN).
SQ SEQUENCE 105 AA; 11484 MW; B427513272E8663D CRC64;

```

```

Query Match
Best Local Similarity 32.7%; Score 153; DB 1; Length 105;
Matches 32; Conservative 17; Mismatches 45; Indels 4; Gaps 2;

```

```

QY 525 POVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQP-ENNYKTPPLVDSGSP 583
DB 6 PEVILPEPSEELKDKATLVCLISDFYRIVKVMKADGNSVTQGVDTTQSSKSNKX 65
QY 584 FLYSKLTVDKSRMVGQGVSCSMVHEALHNHYTQKSLS 621
DB 66 AASFHLTANQWKSYSVTQCVTHB---GHTVEKSLA 100

```

```

RESULT 80
KCB RABIT STANDARD; PRT; 106 AA.
AC P01839;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE IG_kappa-B4 chain C region.
GN K-BAS.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
CC NCBI_TaxID=9986;
XX [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Basilea;
RA Heidmann O., Rougeon P.;
RT "Multiplicity of constant kappa light chain genes in the rabbit
RT genome: a b4b4 homozygous rabbit contains a kappa-bas gene."
RL EMBO J. 2;437-441(1983).
CC -1- MISCELLANEOUS: In Basilea rabbits, the major type of light chain
CC is lambda. The kappa chain shown is a minor component. All other
CC rabbit B allotypes have Cys-64
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

```

CC EMBL; V01241; CAA24558.1;
DR EMBL; V00885; -; NOT_ANNOTATED_CDS.
DR PIR; A02121; K4RBS.
DR HSSP; P01842; 7FAB.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig1.
DR SMART; SM00407; IgC1; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; FALSE NEG.
KW Immunoglobulin domain; Immunoglobulin C region.
KW NON_TER
FT DOMAIN 1 99 IG-LIKE.
FT DISULFID 27 87
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
SQ SEQUENCE 106 AA; 11279 MW; AF9B928DDA853849 CRC64;

```

```

Query Match
Best Local Similarity 33.7%; Score 153; DB 1; Length 106;
Matches 31; Conservative 19; Mismatches 36; Indels 6; Gaps 2;

```

```

QY 520 GQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQP-ENNYKTPP 575
DB 1 GDPVAPSVLTPPPSKELTGTATIVCANRFPYPSDIAVTKVKVDTTQSGIENSKT--P 58
QY 576 VLDSDGPFYLSKLTVDKSRMVGQGVSCSMV 607
DB 59 QSPEDNTYLSSTJLSTLSAQTNSHSHVYTCVAV 90

```

```

RESULT 81
VCAM1 MOUSE STANDARD; PRT; 739 AA.
AC P29533;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vascular cell adhesion protein 1 precursor (V-CAM 1).
GN VCAM1 OR VCAM-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC NCBI_TaxID=10090;
XX [1]
RN SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=FVB; TISSUE=Lung;
RX MEDLINE=92181437; PubMed=1371918;
RX Heselin C., Moy P., Tizard R., Chisholm P., Williams C., Wyke M.,
RX Burdly L., Miyake K., Kincaide P., Lobb R.;
RT "Cloning of murine and rat vascular cell adhesion molecule-1."
RT Biochem. Biophys. Res. Commun. 183;163-169(1992).
XX [2]
RN SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Lymph node;
RX MEDLINE=93246254; PubMed=7683304;
RX Araki M., Araki K., Vassalli P.;
RT "Cloning and sequencing of mouse VCAM-1 cDNA."
RL Gene 126;261-264(1993).
XX [3]
RN SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=129; TISSUE=Embryo;
RX MEDLINE=94117008; PubMed=7507076;
RX Cybulsky M.I., Allan-McCormack M., Collins T.;
RT "Structure of the murine VCAM1 gene."
RL Genomics 18;387-391(1993).
XX [4]
RN SEQUENCE OF 1-693 FROM N.A. (ISOFORM 1).

```


Qy	89	QGNRPFLIKNKLKEDSDTYICE-----VEDQKEVOLV-----FGJLANSDBTHLQOGS	138
Db	271	SGNATLTLMLMRHEDSGVYVCEGNVLIGDKRAVELVVOEKRFYVDVSPGQVAAYQVDS	330
Qy	139	LPTLLESPSSPVQCRSPRCNKIOG-----GKTLSVQLLEIOPSGTWTCTVLONOK	191
Db	331	VLTCAIIGCDSSFSFMRQTOBSPINGVVRNMGASTLVLSVGFEDBHSITLCANVTQOR	390
Qy	192	KVEPKIDIVVLAFOKASIVYKKEGQVEFSPLA-----FTVEKL	232
Db	391	TLEKRTQVEVYSF-----PEDVIMKSGPLVHGRRPTVNTCTPVNVYFPDLHRIEL	441
Qy	233	TGSGELMWOHERKSSSSKSMI-TPDLKNEKVSXKRTQOPKQ-NGKUL-----PLHLTPQ	286
Db	442	KG-----ETTLMKRYFLEEMGSISETKLETTFTPTIEDYKSLVCLARHSGEME	493
Qy	287	ALPOYAGSGNLTALAEAKGKTLHOEYNVLVMMATOLQ-----NLTCBWMGPTSPKMLSL	342
Db	494	SEPKQROSVQPLVYNVAP-----KETTIVWSPLLEEGSPVNLCTSSDGIPAKXILMSR	548
Qy	343	KLENKEAKVSKREKPVWVLN---PEAGWMOCL-LSDSQVILLESNIKVLPWTSTPVEPKS	398
Db	549	QLNNGELQPSLEBNTTLTPMSTKRDSGIVVCGINEAG--ISRSVELI-----IQVSPKD	602
Qy	399	CDKHTHCPCPAPELLGSPSVLFPP--KPKDTLMISRT-----PEVTCV	442
Db	603	IOLT-----VFPEKSVYEGDVTYIISCTGQANVEFTMIL	635

ID	KACA_RAT	STANDARD;	PRT;	106 AA.
AC	P01836;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	IG kappa chain C region, A allele.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=DA;			
RX	MEDLINE=82082587; PubMed=62733908;			
RA	Sheppard H.W., Gorman G.A.;			
RT	"Allelic forms of rat kappa chain genes: evidence for strong			
RT	selection at the level of nucleotide sequence."			
CC	Proc. Natl. Acad. Sci. U.S.A. 78:7064-7068(1981).			
CC	1- SMILARITY: Contains 1 immunoglobulin-like domain.			
DR	PIR: A02118; KIRTA.			
DR	HSSP: P01842; 2MCG.			
DR	InterPro: IPR007110; Ig-1like.			
DR	InterPro: IPR003597; Ig_C1.			
DR	InterPro: IPR003066; IG_MHC.			
DR	Pfam: PF00047; Ig; 1.			
DR	SMART: SM00407; IGc1; 1.			
DR	PROSITE: PS50835; IG_LIKE; 1.			
DR	PROSITE: PS00290; IG_MHC; 1.			
KW	immunoglobulin domain; Immunoglobulin C region.			
FT	NOV TERM	1		
FT	DOMAIN	5	102	IG-LIKE.
FT	DISULFID	26	86	
FT	DISULFID	106	106	INTERCHAIN (WITH A HEAVY CHAIN).
SO	SEQUENCE	106 AA;	11732 MW;	B7E120D9700DD66 CRC64.
QY	Query March	4.4;	Score 150;	DB 1; Length 106;
	Best Local Similarity	31.5%;	Pred. NO 0.0027;	
	Matches 29;	Conservative 23;	Mismatches 28;	Indels 12; Gaps 2
DB	5	PVT5IFPPSASEQELTSCGATVCFVNFFPRDISVSKKIDGSEQRD-----GLDSVTVDQ	59	
DB	5799	525 PQVYTLPPSDELTKQVSLTLCYKGYPSDIAVWESNQPENNYKTTPEVLD-----		

```
Oy      580 --DGSPFLYSKLTVDKSRMOQGNVSCSWME    609  
         |::|::|::|::|::|:  
Db      60 SKDSTYMSSTSLTKVEYERHNLTCCEVVHK   91
```

RESULT 83

ID	SHS1_RAT	STANDARD;	PRT;	509 AA.
AC	P97710;	008951;	070426;	Q9QW15;
DT	10-OCT-2003	(Rel. 42, Created)		
DT	10-OCT-2003	(Rel. 42, Last annotation update)		
DT	10-OCT-2003	(Rel. 42, Last annotation update)		
DE	Protein-tyrosine phosphatase non-receptor type substrate 1 precursor (SHP substrate-1) (SHPS-1) (inhibitory receptor SHPS-1) (Signal-regulatory protein alpha-1) (Src-1-alpha-1) (Bran Ig-like molecule with tyrosine-based activation motifs) (Blt) (Macrophage fusion receptor) (Macrophage membrane protein MFP150).			
GN	PTPNS1 OR SHPS1 OR SIRP OR BIT OR MFR.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euteheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_Taxid:101116;			
RN	[1]			
RP	SEQUENCE FROM N.A., SEQUENCE OF 42-60; 68-91; 128-137; 150-158;			
RP	174-189; 192-202; 204-212; 218-237; 259-270; 278-282; 405-415 AND			
RP	446-453, N-GLYCOSYLATION, PHOSPHORYLATION ON TYROSINE RESIDUES, AND			
RP	INTERACTIONS WITH PTPN6 AND PTPN11.			
RC	TISSUE=Fetal fibroblast;			
RX	MEDLINE=97098667; Pubmed=8943344;			
RA	Fujioka Y., Macozaki T., Noguchi T., Iwamatsu A., Yamao T.,			
RA	Takahashi N., Tauda M., Takada T., Kasuga M.;			
RT	"A novel membrane glycoprotein, SHS-1, that binds the SH2 domain-			
RT	containing protein tyrosine phosphatase SHP-2 in response to mitogens			
RT	and cell adhesion.";			
RL	Mol. Cell. Biol. 16:6887-6899(1996).			
RN	[2]			
RP	SEQUENCE FROM N.A., SEQUENCE OF 32-48 AND 446-453, FUNCTION, AND			
RP	PHOSPHORYLATION ON TYROSINE RESIDUES.			
RC	STRAIN=Sprague-Dawley; TISSUE=Brain.			
RX	MEDLINE=97415431; Pubmed=9271230;			
RA	Sano S.-I., Ohnishi H., Omori A., Hasegawa J., Kubota M.;			
RT	"BIT, an immune antigen receptor-like molecule in the brain.";			
RL	FEBS Lett. 411:327-334(1997).			
RN	[3]			
RP	SEQUENCE FROM N.A., SEQUENCE OF 99-107; 128-149; 192-217; 405-417;			
RP	419-429; 446-467 AND 496-506, N-GLYCOSYLATION, AND TISSUE SPECIFICITY			
RC	STRAIN=Flischer 344; TISSUE=Macrophage;			
RX	MEDLINE=98449911; Pubmed=9776638;			
RA	Sadeghifar C., Sterling H., Beckers C., Kobayashi R., Solimena M.,			
RA	Ullu E., Vignery A.;			
RT	"MFR, a putative receptor mediating the fusion of macrophages.";			
RL	Mol. Cell. Biol. 18:6213-6223(1998).			
RN	[4]			
RP	SEQUENCE OF 1-419 FROM N.A., AND TISSUE SPECIFICITY.			
RC	STRAIN=MAG/Rij; TISSUE=Alveolar macrophage;			
RX	MEDLINE=98375871; Pubmed=9712053;			
RA	Adams S., van der Laan L.J.W., Vernon-Wilson E.,			
RA	Renardel de Lavalette C., Doepp E.A., Dijkstra C.D., Simons D.L.,			
RA	van den Berg T.K.;			
RT	"Signal-regulatory protein is selectively expressed by myeloid and			
RT	neural cells.";			
RL	J. Immunol. 161:1853-1859(1998).			
RN	[5]			
RP	PHOSPHORYLATION IN RESPONSE TO EGF, AND INTERACTION WITH PTPN11.			
RX	MEDLINE=98008865; Pubmed=9344855;			
RA	Ouchi F., Macozaki T., Noguchi T., Fujioka Y., Yamao T., Takada T.,			
RA	Tsuda M., Takeda H., Fukunaga K., Okabayashi Y., Kasuga M.;			
RT	"Epidermal growth factor stimulates the tyrosine phosphorylation of			
RT	SHS-1 and association of SHS-1 with SHP-2, a SH2 domain-containing			
RT	protein tyrosine phosphatase.";			
RL	Biochem. Biophys. Res. Commun. 239:483-487(1997).			
RN	[6]			

RP	PHOSPHORYLATION OF TYR-477 AND TYR-501, AND MUTAGENESIS OF TYR-436;
RP	TYR-460; TYR-477 AND TYR-501.
RK	MEDLINE=968204923; PubMed=9535915;
RA	Takeda T., Matozaki T., Takeda H., Fukunaga K., Noguchi T.,
RA	Fujioka Y., Okazaki I., Tsuda M., Yamao T., Ochi F., Kasuga M.;
RT	"Roles of the complex formation of SHS-1 with SHP-2 in insulin-
RT	insulin-stimulated mitogen-activated protein kinase activation.";
RL	J. Biol. Chem. 273:9234-9242(1998).
CC	-1- PUNCTION: Immunoglobulin-like cell surface receptor for CD47. Acts
CC	as docking protein and induces translocation of PTPN6,
CC	PTPN11 and other binding partners from the cytosol to the
CC	plasma membrane. Supports adhesion of cerebellar neurons, neurite
CC	outgrowth and glial cell attachment. May play a key role in
CC	intracellular signaling during synaptogenesis and in synaptic
CC	function. INVOLVED in the negative regulation of receptor tyrosine
CC	kinase-coupled cellular responses induced by cell adhesion, growth
CC	factors or insulin. Mediates negative regulation of phagocytosis,
CC	mast cell activation and dendritic cell activation. CD47 binding
CC	prevents maturation of immature dendritic cells and inhibits
CC	cytokine production by mature dendritic cells. May play a role in
CC	the release of nitric oxide by macrophages (By similarity).
CC	-1- SUBUNIT: Binds PTPN11 when tyrosine-phosphorylated, except in
CC	macrophages, where it primarily binds PTPN6. Binds GRB2 in vitro.
CC	Binds FGR. Binds JAK2 irrespective of its phosphorylation status
CC	and forms a stable complex. Binds SCAPI and/or SCAP2. The
CC	resulting complex recruits FTY. Binds PRK28 (By similarity).
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.
CC	-1- TISSUE SPECIFICITY: Highly expressed in brain, spleen, lung, liver
CC	and kidney. Detected at lower levels in heart. Highly expressed in
CC	alveolar and peritoneal macrophages, and at lower levels in
CC	dendritic cells.
CC	-1- PTM: N-glycosylated.
CC	-1- PTM: Phosphorylated on tyrosine residues in response to insulin,
CC	cell adhesion or epidermal growth factors. Dephosphorylated by
CC	PTPN11.
CC	-1- SIMILARITY: Contains 2 immunoglobulin-like C1-type domains.
CC	-1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-sb.ch/announce
CC	or send an email to license@isb-sb.ch).
CC	-----
DR	EMBL; D85183; BAA12734.1; -;
DR	EMBL; D38468; BAA20368.1; -;
DR	EMBL; U62328; AAC68478.1; -;
DR	EMBL; AF055065; AAC18089.1; -;
DR	HSSP; P01703; 7FAB.
DR	InterPro; IPR007110; IG-1-like.
DR	InterPro; IPR003597; IG_C1.
DR	InterPro; IPR003006; IG_MHC.
DR	Pfam; PF00047; Ig_3.
DR	SMART; SM00407; IgC1. 2.
DR	PROSITE; PS50835; IG_LIKE; 3.
DR	PROSITE; PS00290; IG_MHC; 1.
KM	Repeat: Signal; Transmembrane; Immunoglobulin domain; SH3-binding;
KM	Glycoprotein; Phosphorylation.
FT	SIGNAL 1 31
FT	CHAIN 32 509
FT	-----
FT	DOMAIN 28 373
FT	TRANSEM 374 394
FT	DOMAIN 395 509
FT	DOMAIN 32 138
FT	DOMAIN 150 248
FT	DOMAIN 255 349
FT	DISULFID 55 122
FT	DISULFID 172 229
FT	DISULFID 274 332
FT	SITE 436 439
FT	-----
FT	SH2-BINDING (POTENTIAL).
FT	-----
FT	PROTEIN-TYROSINE PHOSPHATASE NON-RECEPTOR
FT	TYPE SUBSTRATE 1.
FT	EXTRACELLULAR (POTENTIAL).
FT	POTENTIAL.
FT	CYTOPLASMIC (POTENTIAL).
FT	IG-LIKE V-TYPE.
FT	IG-LIKE C1-TYPE 1.
FT	IG-LIKE C1-TYPE 2.
FT	POTENTIAL.
FT	POTENTIAL.
FT	POTENTIAL.
FT	SH2-BINDING (POTENTIAL).

Query Match	Similarity	Score	DB 1	Length	509
Best Local	Similarity	21.6%	Pred. No.	0.021	
Matches	Conservative	63	Mismatches	158	Indels 80; Gaps 16
259	KEYSVKRYTQDPKLGKGLPLHLTLPQALP-----QVAGSNLTLLAEAKTGKLGQVNV	313			
32	KELVATQADKSVSVAAGDSATLNTCTVSLTFVGPDKWKGSGQNRSPYISFGSHFPRI	91			
314	LVMRATQDLOGNLTCVWVGPTSPKLMSTLKENNEAKVSKREKRPVWVNLNPAAGMOCILLS	373			
92	TNVSADATK-RKNMD-----FSICISN-----VTPEDGITYCYKVF	125			
374	DSGCVLLESNIK-----VLPMTWSTFVEBPKSCDKHTTCCPCAPPELLGSGSVFLFPK	425			
126	QKGIIVEPPTTEIKSGGGTLLVYL-----AKRSS-----FEVSGPDSRGSP-----	164			
426	PKDTLMISRTPEVTVGVVVDVSHEDPEVKVFNMYVDGVEVHNAKTKRREQVNSTYRVSVYL	485			


```

Db 165 -----GQTVNFTCKSYGFSPPN--ITLKLKDGKELSHLETTISKS-NVSYNISSTV 214
Qy 486 TV-LHODMNGKXKCVSNKALPA-PIEKTISKAGQPREPOVYTLPPSRDEL-T-KNOV 542
Db 215 SYKLSEEDHSRVI-CEVAHVTLGRLNGTANFSIIRVSP---TLTKTQOPLTPASGV 270
Qy 543 SLTCLVKGFPSPDIWESNGCPENNYKTPPVLDSDGSFFLYSKLTVDKSRMOGNVF 602
Db 271 NLTCQVQKFPKALQNLMLNGLSRDKEPHFTDRDGTGVNTSLFLVNSSAHREDYVF 330
Qy 603 SCGVME---ALNHTYTKSLSLIS 623
Db 331 TCVEHDSQPAITENTHTRAFHNS 354

RESULT 84
HB2C_PIG STANDARD; PRT; 261 AA.
ID HB2C_PIG
AC P15982;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE SLA class II histocompatibility antigen, DQ haplotype C beta chain
DE precursor.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN (1)
RP SEQUENCE FROM N.A.
RA MEDLINE=90361905; PubMed=2391424;
RA Gustafsson K., Leguenn C., Hirsch F., Germana S., Pratt K.,
RA Sachs D.H.;
RT "Class II genes of miniature swine. IV. Characterization and
RT expression of two allelic class II DQB cDNA clones.";
RL J. Immunol. 145:1946-1951(1990).
RC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M31497; AAA1084.1; -
DR EMBL: M32117; AAA53110.1; -
DR PIR: A60404; A60404.
DR HSSP: P13760; 2SEB.
DR InterPro: IPR007110; IG_1like.
DR InterPro: IPR003597; IG_c1.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR000353; MHC_II_beta.
DR pfam: PFO0047; IG_1.
DR pfam: PFO0069; MHC_II_beta; 1.
DR ProDom: PD000328; MHC_II_beta; 1.
DR SMART: SM00407; IGc1_1.
DR PROSITE: PSS0835; IG_1like; 1.
DR PROSITE: PS00290; IG_MHC; 1.
DR MHC II; Transmembrane; Glycoprotein; Signal.
KW SIGNAL
FT CHAIN 1 31
FT 32 261 SLA CLASS II HISTOCOMPATIBILITY ANTIGEN,
FT DO HAPLOTYPE C BETA CHAIN.
FT DOMAIN 32 126
FT DOMAIN 127 220 EXTRAEXONULAR BETA-1.
FT DOMAIN 221 230 EXTRAEXONULAR BETA-2.
FT TRANSMEM 231 251 CONNECTING PEPTIDE.
FT DOMAIN 252 261
FT DOMAIN 261 261 CYTOPLASMIC TAIL.
FT DISULFID 47 111 BY SIMILARITY.
FT DISULFID 149 205 BY SIMILARITY.
FT CARBOHYD 51 51 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 261 AA; 29550 MW; AA9581F2A3B1969D CRC64;

```

```

Query Match 4.4%; Score 149; DB 1; Length 261;
Best Local Similarity 24.8%; Pred. No. 0.01;
Matches 51; Conservative 36; Mismatches 99; Indels 20; Gaps 6;

Qy 409 PAPELLGSPVFLPPPKKDTLMISRPEVTCVVDVDSHEDPEVKFNMYVDGVEYHNAKT 468
Db 28 PVAERDRSPDFVFOFK-GECEFYNGTQVRGVARYIYNQEHILRFDSDVEFRAVTPLG 86
Qy 469 KPREQVSTTRVYVVLTVLHODMNGKXKCVSNKALPA-PIEKTISKAGQPREPOVY 528
Db 87 RPEADSNWSQKDVLEQMPA-EYDRCVKNYQLE-----ESTTLQRRVQ---TV 131
Qy 529 TLPPSRDELITKXQVSLTCLVKGFPSPDIWESNGCPEN-NYKTPPVLDSDGSFFLYS 587
Db 132 TISPESKALALHNNLVCALVDFYPSQYKQVQFNRGQETGAGVSTPLIRNGD---WYYQ 188
Qy 588 KLTVDKSRMOQGNVFCGVMEBALHN 613
Db 189 VLVMLENNLORGDDVYTCRVEHSLQN 214

RESULT 85
DSCA_HUMAN STANDARD; PRT; 2012 AA.
ID DSCA_HUMAN
AC O60469; O60468;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Down syndrome cell adhesion molecule precursor (CHD2).
DE DSCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Brain;
RX MEDLINE=98087574; PubMed=9426258;
RA Yamakawa K., Huot Y.-K., Haendelt M.A., Hubert R., Chen X.-N.,
RA Lyons G.E., Korenberg J.R.;
RT "DSCAM: a novel member of the immunoglobulin superfamily maps in a
RT Down syndrome region and is involved in the development of the
RT nervous system.";
RL Hum. Mol. Genet. 7:227-237(1998).
RN (2)
RP SEQUENCE FROM N.A., AND FUNCTION.
RX MEDLINE=20384934; PubMed=10925149;
RA Agarwala K.L., Nakamura S., Teutsuni Y., Yamakawa K.;
RT "Down syndrome cell adhesion molecule DSCAM mediates homophilic
RT intercellular adhesion.";
RL Brain Res. Mol. Brain Res. 79:118-126(2000).
RN (3)
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totsuki Y., Choi D.-K., Groner Y.,
RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyma S., Antonarakis S.E.,
RA Mitsuima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecher H.,
RA Ramer J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21. ";
RL Nature 405:311-319(2000).
CC -I- FUNCTION: CELL ADHESION MOLECULE THAT CAN MEDIATE CATION-
CC INDEPENDENT HOMOPHILIC BINDING ACTIVITY. COULD BE INVOLVED IN
CC NERVOUS SYSTEM DEVELOPMENT.
CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PROBABLE). THE
CC SHORT ISOFORM MAY BE SECRETED.

```

```

CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long; Synonyms=CHD2-52;
CC IsoId=O60469-1; Sequence=Displayed;
CC Name=Short; Synonyms=CHD2-42;
CC IsoId=O60469-2; Sequence=VSP_002502, VSP_002503;
CC -1- TISSUE SPECIFICITY: Primarily expressed in brain.
CC -1- SIMILARITY: Contains 10 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 6 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF023450; AAC17967.1; -
CC EMBL; AF023449; AAC17966.1; -
CC EMBL; AF217525; AAF27525.1; -
CC EMBL; AL163283; CAB90464.1; -
CC EMBL; AL163282; CAB90435.1; -
CC EMBL; AL163281; CAB90444.1; -
CC Genew; HGNC:3039; DSCAM.
CC MIM; 602523; -
CC DR GO; GO:0005887; C: integral to plasma membrane; TAS.
CC DR GO; GO:0005624; C: membrane fraction; TAS.
CC DR GO; GO:0007155; P: cell adhesion; TAS.
CC DR GO; GO:0007399; P: neurogenesis; TAS.
CC DR InterPro; IPR003957; FN_III-like.
CC DR InterPro; IPR007110; IG_III-like.
CC DR InterPro; IPR003598; IG_c2.
CC Pfam; PF00041; fn3; 6.
CC Pfam; PF00047; igf; 9.
CC SMART; SM00060; FN3; 6.
CC SMART; SM00408; IGc2; 7.
CC PROSITE; PS00835; IG_LIKE; 9.
CC DR Immunoglobulin domain; Glycoprotein; signal; Cell adhesion; Repeat;
CC Transmembrane; Alternative splicing.
CC KW SIGNAL.
CC FT CHAIN 1 17 POTENTIAL.
CC FT CHAIN 18 2012 DOWN SYNDROME CELL ADHESION MOLECULE.
CC FT TRANSMEM 1596 1616 EXTRACELLULAR (POTENTIAL).
CC FT DOMAIN 1617 2012 POTENTIAL.
CC FT DOMAIN 125 129 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 125 216 IG-LIKE C2-TYPE 1.
CC FT DOMAIN 225 305 IG-LIKE C2-TYPE 2.
CC FT DOMAIN 313 401 IG-LIKE C2-TYPE 3.
CC FT DOMAIN 407 500 IG-LIKE C2-TYPE 4.
CC FT DOMAIN 504 592 IG-LIKE C2-TYPE 5.
CC FT DOMAIN 596 685 IG-LIKE C2-TYPE 6.
CC FT DOMAIN 690 783 IG-LIKE C2-TYPE 7.
CC FT DOMAIN 787 883 IG-LIKE C2-TYPE 8.
CC FT DOMAIN 885 972 IG-LIKE C2-TYPE 9.
CC FT DOMAIN 972 1076 FIBRONECTIN TYPE-III 1.
CC FT DOMAIN 1088 1177 FIBRONECTIN TYPE-III 2.
CC FT DOMAIN 1189 1273 FIBRONECTIN TYPE-III 3.
CC FT DOMAIN 1285 1377 FIBRONECTIN TYPE-III 4.
CC FT DOMAIN 1380 1463 FIBRONECTIN TYPE-III 5.
CC FT DOMAIN 1477 1562 FIBRONECTIN TYPE-III 6.
CC FT DISULFID 46 102 BY SIMILARITY.
CC FT DISULFID 145 197 BY SIMILARITY.
CC FT DISULFID 246 293 BY SIMILARITY.
CC FT DISULFID 335 385 BY SIMILARITY.
CC FT DISULFID 428 484 BY SIMILARITY.
CC FT DISULFID 525 575 BY SIMILARITY.
CC FT DISULFID 617 669 BY SIMILARITY.
CC FT DISULFID 711 766 BY SIMILARITY.
CC FT DISULFID 809 865 BY SIMILARITY.
CC FT DISULFID 1307 1359 BY SIMILARITY.
CC FT CARBOHYD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 487 487 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 512 512 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 658 658 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 666 666 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 748 748 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 795 795 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 924 924 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1142 1142 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1160 1160 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1250 1250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1271 1271 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1341 1341 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1488 1488 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1562 1571 N-GLYCOSYLATION SITE (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1572 2012 /FtId=VSP_002502.
FT VARSPLIC 1893 2012 Missing (in isoform Short).
FT CONFLICT 1893 2012 /FtId=VSP_002503.
FT CONFLICT 1893 2012 HRRDILHLPYILMDPLNRGGRGTRDLSGACLEPOK
FT FT SRTLKRPVLEPIEMEAASSASTREGOSQOPGAVLPOR
FT FT EGAEIQAAMSSQESLDSRGLKGNPNYAKSYTLV ->
FT FT IGVTSYICHTLEWFC (IN REF. 1).
FT SO SEQUENCE 2012 AA; 222259 MW; 0E33CFB781A08334 CRC64;

Query Match 4.3% Score 148.5; DB 1; Length 2012;
Best Local Similarity 20.3%; Pred. No. 0.16;
Matches 130; Conservative 78; Mismatches 230; Indels 203; Gaps 32;

34 GDTVELTCTASOKKSIQFHKNKSNQIKILGNQSGFLTKGSKLNDRAISRRLDQGNFP 93
239 GQRYELCKLGHPEPRYLKDNM-----PLELSGR-----FQKVTG 277
94 LIINKLKIEDSDTYICEVEDQKEVOLVFLGFLANSPTHLLOQSLTTLT-----ESPFG 148
278 LIENIRPDSGSGVCEVSNR-----YG-TAKVIGRLVYVQPLKATISPRKVSVG 328
149 SSPEVQR-----SPRKNIO-----GKTLSSVSLQLEODSGTWICTV 186
329 SQVSLSCSVTGTEDQLSWYRNGEILNP-GKNVITGINHEHILMDHVSXDGAYOCFV 387
187 LONQKTEFKIDYVLAFOKASSIVYKEGEQVFSPLFTVEKLTGSGELMWAQBRAS 246
388 RKDLSAQDYVQVL-----EDTPKILS--AF-----SKRV 418
247 SSKSWITFDLKNKEVSVKVT---ODPKLQMK-----KLPLHLTLPLQALPOY 291
419 SPAEPVSLMCNVKGTPLPLITLTLDDPILKGSHRISQMTSEGNVSYLNISSQVDR 478
292 AGSNLTLAEATGKGLHDEVNLVWRKATOLQKLT-----CEV-----NG 332
479 GGVYRCTANNSAGVLYQARINVRGPASIRPMKNITIAGRDTYIHGRVIGYPYSIKWY 538
333 PTPSKMLSLKL-----ENKEAKYSKEKPYWNLNPEAGMOCLDSGOYLSENIKLP 388
539 KNSLPLFPNRQVAFENNGTLKLSDVQKEV-----DEGEVTC-----NLVQF 581
389 TWSTPVPKSCDKTHTCPPCAPPELLGSPVFLFPPPKDQTLTMSIRTEVTCVVVDVSH 448
582 QLSI---SQSVHTVVKVPFIQF-----FEPP-----RFSIGQVFIQCVVVS---G 622
449 DPEVKFMYVD-----GVEVNAKTKRREQYNSYTRVSVLYTLHODMLNGKRYKC 500
623 DLPTITWQDGRPIPSGLVTIDNI-----DFTSSLR-ISLTLIMH-----NG-NYTC 669
501 KVSNAKLPARIEKTISSAKGQPREPOVYTLPSRDELTKQVQSITCVKQYPSDIADVW 560
670 IARNEA--AAVEHQSLIVAVP--PKRVVQPRQDGIYGAAVILNCSAEG-YFPTIYWK 724
561 ESNQ-----QP-----ENNYKTPPVLDSDGSFPL 585

```

Db 725 FSKGAGVQFQFIALNGRIQVLSNGSLIKHVEEDSGYLL 765

RESULT 86
KAC_HUMAN STANDARD; PRT; 106 AA.
AC P01834;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 1g kappa chain C region.
GN IGKC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCB1_Taxid9606;
RN (1)
RP SEQUENCE (MYELOMA PROTEIN EU).
RX MEDLINE=71064023; PubMed=5489770;
RA Gottlieb P.D., Cunningham B.A., Rutishauser U., Edelman G.M.;
RT "The covalent structure of a human gamma G-immunoglobulin. VI. Amino
RT acid sequence of the light chain.";
RL Biochemistry 9:3155-3161(1970).
RN (2)
RP DISULFIDE BONDS.
RX MEDLINE=71064027; PubMed=4923144;
RA Calli W.E., Edelman G.M.;
RT "The covalent structure of a human gamma G-immunoglobulin. X.
RT intrachain disulfide bonds.";
RL Biochemistry 9:3188-3196(1970).
RN (3)
RP SEQUENCE (BENCE-JONES PROTEIN TI).
RX MEDLINE=72188439; PubMed=50277703;
RA Suter L., Barnikol H.U., Watanabe S., Hilschmann N.;
RT "Rule of antibody structure. The primary structure of a monoclonal
RT immunoglobulin L-chain of kappa-type, subgroup 3 (Bence-Jones protein
RT TI). IV. The complete amino acid sequence and its significance for
RT the mechanism of antibody production.";
RL Hoppe-Seyler's Z. Physiol. Chem. 353:189-208(1972).
RN (4)
RP SEQUENCE FROM N.A.
RX MEDLINE=81042304; PubMed=6775818;
RA Hleber P.A., Max E.E., Seidman J.G., Maizel J.V. Jr., Leder P.;
RT "Cloned human and mouse kappa immunoglobulin constant and J region
RT gene conserve homology in functional segments.";
RL Cell 22:197-207(1980).
RN (5)
RP SEQUENCE (BENCE-JONES PROTEIN ROY).
RA Hilschmann N., Barnikol H.U., Hess M., Langer B., Ponertingl H.,
RL Steinmetz-Kayne M., Suter L., Watanabe S.;
RT (in) Franek F., Shugar D. (eds.);
RL Gamma globulins: structure and function, pp.57-74, Academic Press,
RL New York (1969).
RN (6)
RP SEQUENCE (BENCE-JONES PROTEIN CUM).
RX MEDLINE=68242259; PubMed=5586923;
RA Hilschmann N.;
RT "The complete amino acid sequence of Bence Jones protein Cum (kappa-
RT type).";
RL Hoppe-Seyler's Z. Physiol. Chem. 348:1718-1722(1967).
RN (7)
RP SEQUENCE (BENCE-JONES PROTEIN AG).
RX MEDLINE=69234743; PubMed=4893682;
RA Titani K., Shinoda T., Putnam F.W.;
RT "The amino acid sequence of a kappa type Bence-Jones protein. 3. The
RT complete sequence and the location of the disulfide bridges.";
RL J. Biol. Chem. 244:3550-3560(1969).
RN (8)
RP SEQUENCE (WALDENSTROM'S MACROGLOBULIN OU).
RX MEDLINE=70201507; PubMed=5447531;
RA Kohler H., Shimizu A., Paul C., Putnam F.W.;
RT "Macroglobulin structure: variable sequence of light and heavy

RT chains.";
RL Science 169:56-59(1970).
RN (9)
RP SEQUENCE OF 1-33; 38-41 AND 62-80.
RC TISSUE=abdominal adipose tissue;
RX MEDLINE=98249779; PubMed=9588180;
RA Olsen K.E., Sletten K., Westermarck P.;
RT "Extended analysis of AL-amyloid protein from abdominal wall
RT subcutaneous fat biopsy: kappa IV immunoglobulin light chain.";
RL Biochem. Biophys. Res. Commun. 245:713-716(1998).
CC -I- MISCELLANEOUS: The EU sequence has the INV (3) allotypic marker,
CC 45-Ala and 83-Val. The ROY sequence has the INV (1,2) allotypic
CC marker, 45-Ala and 83-Leu.
CC -I- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J00241; AAA58989.1; -.
DR EMBL; V00557; CAA23823.1; -.
DR PIR; B90562; K3HU.
DR PDB; 1D5B; 09-FEB-00.
DR PDB; 1D5I; 09-FEB-00.
DR PDB; 1D6V; 04-OCT-00.
DR PDB; 1HE2; 18-JUL-03.
DR PDB; 1HKL; 12-MAR-97.
DR PDB; 1I72; 08-AUG-01.
DR PDB; 1M1M; 15-MAY-97.
DR Genew; HGNC:5716; IGKC.
DR MIM; 147200; -.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG-cl.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF000407; IGc1; 1.
DR SMART; SM00407; IGc1; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KM Immunoglobulin domain; Immunoglobulin C region; 3D-structure.
FT NON_TER 1 1
FT DOMAIN 5 102 IG-LIKE.
FT DISULFID 26 86 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 106 106 V -> L (IN INV(1,2) MARKER).
FT VARIANT 83 83 /FTID=VAR_003897.
FT CONFLICT 14 14 D -> N (IN REF. 5 AND 6).
FT CONFLICT 57 57 E -> Q (IN REF. 5 AND 6).
SQ SEQUENCE 106 AA; 11609 MW; 51984D1FDD372CE8 CRC64;

Query Match 4.3%; Score 148; DB 1; Length 106;
Best Local Similarity 29.5%; Pred. No. 0.0036;
Matches 28; Conservative 24; Mismatches 41; Indels 2; Gaps 1;

Qy 525 POUYTLPPSRDELTLNGQVSLTCLVGFYPSDIAVEMESNG--OPENNYKTPPVLDSPGS 582
Db 5 PSVFIFPSDEQLSGTASVCLNFTYPRAKYQMKDNLQSGNSGVTEQDSKST 64

Qy 563 FFLYSKLTVDYKSRMQGQNVFSCSVMEALAHNYQ 617
Db 65 YLSSTLTLSKADYKHKVYACEVTHQGLSPVTK 99

RESULT 87
DCC_MOUSE STANDARD; PRT; 1447 AA.
AC P70211;
DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Tumor suppressor protein DCC precursor.
 GN DCC.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS A; B AND C).
 RC STRAIN=BALB/c; TISSUE=Brain;
 RX MEDLINE=96112625; PubMed=8570174;
 RA Cooper H.W., Ames P., Brito J., Gad J., Wilks A.F.;
 RT "Cloning of the mouse homologue of the deleted in colorectal cancer
 gene (mdcc) and its expression in the developing mouse embryo.";
 RL Oncogene 11:2243-2254(1995).
 RP REVISIONS.
 RC STRAIN=BALB/c; TISSUE=Brain;
 RA Cooper H.W.;
 RL Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=A;
 CC IsoId=P70211-1; Sequence=Displayed;
 CC Note=Isoform B is produced by alternative initiation at Met-85
 CC of isoform A;
 CC Name=C;
 CC IsoId=P70211-2; Sequence=VSP_002501;
 CC Event=Alternative initiation;
 CC Comment=2 isoforms, A (shown here) and B, are produced by
 CC alternative initiation at Met-1 and Met-85;
 CC -1- TISSUE SPECIFICITY: In the embryo, expressed at high levels in the
 CC developing brain and neural tube. In adult, highly expressed in
 CC brain with very low levels found in testis, heart and thymus.
 CC Isoform C is expressed only in the embryo.
 CC -1- DEVELOPMENTAL STAGE: Low levels in early gestation. Highest levels
 CC expressed during mid gestation. Levels decrease in late gestation
 CC and remain at this level in the adult.
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. DCC family.
 CC -1- SIMILARITY: Contains 4 immunoglobulin-like C2-type domains.
 CC -1- SIMILARITY: Contains 6 fibronectin type III domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X85788; CAA59786.1; -.
 DR HSSP; P56276; ITLK.
 DR MGD; MGI:94869; DCC.
 DR InterPro; IPR008957; FN_III-like.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR003962; FNIII_subd.
 DR InterPro; IPR007110; Ig_IIIc.
 DR InterPro; IPR003598; Ig_C2.
 DR Pfam; PF00041; Ig3; 6.
 DR Pfam; PF00047; Ig3; 4.
 DR PRINTS; PR00014; FNTYPEIII.
 DR SMART; SM00060; FN3; 6.
 DR SMART; SM00408; IGC2; 3.
 DR PROSITE; PS50835; IG_LIKE; 4.
 KM Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
 KM Repeat; Anti-oncogene; Alternative initiation; Alternative splicing.
 FT SIGNAL 1 25
 FT CHAIN 26 1447 TUMOR SUPPRESSOR PROTEIN DCC, ISOFORM A.
 FT CHAIN 85 1447 TUMOR SUPPRESSOR PROTEIN DCC, ISOFORM B.
 FT INIT_MET 85 85 FOR ISOFORM B.

FT DOMAIN 26 1097 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1098 1122 POTENTIAL.
 FT DOMAIN 1123 1447 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 133 135 IG-LIKE C2-TYPE 1.
 FT DOMAIN 139 229 IG-LIKE C2-TYPE 2.
 FT DOMAIN 234 326 IG-LIKE C2-TYPE 3.
 FT DOMAIN 331 416 IG-LIKE C2-TYPE 4.
 FT DOMAIN 426 522 FIBONECTIN TYPE-III 1.
 FT DOMAIN 525 618 FIBONECTIN TYPE-III 2.
 FT DOMAIN 619 716 FIBONECTIN TYPE-III 3.
 FT DOMAIN 722 816 FIBONECTIN TYPE-III 4.
 FT DOMAIN 840 940 FIBONECTIN TYPE-III 5.
 FT DOMAIN 941 1042 FIBONECTIN TYPE-III 6.
 FT DISULFID 61 117 BY SIMILARITY.
 FT DISULFID 161 212 BY SIMILARITY.
 FT DISULFID 261 310 BY SIMILARITY.
 FT DISULFID 352 400 BY SIMILARITY.
 FT CARBOHYD 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 628 628 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 702 702 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 819 838 Missing (in isoform C).
 FT FT
 SQ SEQUENCE 1447 AA; 158298 MW; 0D1F1097C22D5B9F CRC64;
 Query Match 4.3%; Score 148; DB 1; Length 1447;
 Best Local Similarity 18.6%; Pred. No. 0.11;
 Matches 121; Conservative 80; Mismatches 205; Indels 244; Gaps 28;
 QY 34 GDTVELTCTASQKSIQFHKNKSNQIKILGNQSGFLTKGSKLND-RADSRSLMDQGNF 92
 DB 154 GDTVLKCEVIGEMPTIHHQKNQD-----DNLPLPGDSRRVVLPSG-- 195
 QY 93 PLIKIKIKIEDSPYICEVED-----QKEVQLD-----VFGLTANSDFHLLQSQ 137
 DB 196 ALQISRLQPDSDGVRCSANPNASIRTNENAEVILSDPGLHRLQVFLQRPNVIAIEGK 255
 QY 138 SLTL---TLESPPGS-----SESVQCRSPRGKNIQSGKTLVSQLELOSSGTMTCVLYQ 188
 DB 256 DAVLECCVSGVPPSPFTMLGEEVITQIRSK--VSLGGSNLILSNVTDSDGYTTCVTVY 314
 QY 189 NOKKVEFKIDIVLA---FOKASSIVYKKEGEVFEFPLAFVTEKLTGSGELMWQAEARA 245
 DB 315 KNEIISASALTYLVLPWFNLHPSNLYAYESMIDF-----ECA 353
 QY 246 SSSKSWITFD-LNKKEVSVKRVTDPRKLGKKLPLHLTLPLQALPOYAGSGNLTTL----- 299
 DB 354 VSGKPVPTVMMKNGDV-----VIPSDYFQIVGGSNLTILGVVK 392
 QY 300 -----ALEATGKLGHOENVLMWRATQLQKNLTCEVWGSPKMLSLKLENKAKV 351
 DB 393 SDGEPYQCVANENAGNAQSSAQLIVPPALPSSIL-----PSAPRVLLVLVSSRFVRL 447
 QY 352 SKKEKPVWVLPNPAQMWQCLLSDGQVLTESNIKVLPTWSTPEVPSKCDTHTCPCPCAP 411
 DB 448 S-----KPPPAEAKNIOTFT----- 463
 QY 412 ELLGSGVFLFPKPKDPTLMSITRPEVTCVVDVSHEDPEVKENWYDGVVHNAKTKPR 471
 DB 464 -----VF-FSRGDRERALTNTQGSLLQTLGNLKKPEAMV----- 498
 QY 472 EEOYNSITRYVSVLTVLHDMVNGKEYK-CKVSNK---ALPAPEKTISSAKQPREPOV 527
 DB 499 -----TFRVVA-----YNEWGESQPIKVAIQPELQVPGVEBNL----- 534
 QY 528 YTLPPSDELTKNOVSLTCLVKGVPYSDIIVENSGQPNVNYKTPPVLYDSQSPFLYS 587
 DB 535 -----HAVSIS-----PISLITIE-----PPAY-ANGPVQGYR 562
 QY 588 KLTVDKSRMOQGN--VFSGSVMEAL--HNHYTQKSLSLSS---PGLQJDE 630

Db 563 LFCTEVTSGKEQNI EVDGSLGKLEGLKKETREYTLRFLAVRYRGPSTDD 612

RESULT 88

SHS1_HUMAN STANDARD; PRT; 503 AA.

AC P78334; 000683; Q43799; Q8N517; Q8TAL8; Q9H022; Q9UDX2; Q9UIJ6;

AC Q9Y4U9;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DE 10-OCT-2003 (Rel. 42, Last annotation update)

DE Protein-tyrosine phosphatase non-receptor type substructure 1 precursor

DE (SHP substructure-1) (SHPS-1) (inhibitory receptor SHPS-1) (Signal-

DE regulatory protein alpha-1) (SHP-alpha-1) (SHP-alpha-2) (SHP-alpha-

DE 3) (MYD-1 antigen) (Brain Ig-like molecule with tyrosine-based

DE activation motifs) (Blt) (Macrophage fusion receptor) (p84).

GN PTPN1 OR SHPS1 OR SIRP OR MYD1 OR BIT OR MFR.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OC NCBI_TaxID=9606;

RN (1)

RP SEQUENCE FROM N.A. (ISOFORM 1).

RC TISSUE=Brain;

RX MEDLINE=97223399; PubMed=9070220;

RA Yamao T., Matozaki T., Amano K., Matsuda Y., Takahashi N., Ochi F.,

RA Fujioaka Y., Kasuga M.,

RT Mouse and human SHPS-1: molecular cloning of cDNAs and chromosomal

RT localization of genes."

RL Biochem. Biophys. Res. Commun. 231:61-67(1997).

RN (2)

RP SEQUENCE FROM N.A. (ISOFORM 1), POLYMORPHISM, PHOSPHORYLATION,

RP N-GLYCOSYLATION, AND INTERACTIONS WITH PTPN11; PTPN6 AND GRB2.

RC TISSUE=Placenta;

RX MEDLINE=97215901; PubMed=9062191;

RA Khartonenkov A., Chen Z., Sures I., Wang H., Schilling J.,

RA Ullrich A.;

RT "A family of proteins that inhibit signalling through tyrosine kinase

RT receptors."

RL Nature 386:181-186(1997).

RN (3)

RP SEQUENCE FROM N.A. (ISOFORM 3), AND VARIANTS ASP-95; LEU-96; ASN-100;

RP ARG-107; GLY-109 AND VAL-131.

RC TISSUE=Monocytes;

RX MEDLINE=98143722; PubMed=9485180;

RA Brooke G.P., Parsons K.R., Howard C.J.;

RT "Cloning of two members of the SIRP alpha family of protein tyrosine

RT phosphatase binding proteins in cattle that are expressed on monocytes

RT and a subpopulation of dendritic cells and which mediate binding to

RT CD4 T cells."

RL Eur. J. Immunol. 28:1-11(1998).

RN (4)

RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANTS LEU-44; THR-50; THR-52;

RP ARG-54; ALA-57; GLY-75; ASP-95; LEU-96; ASN-100; ARG-107; GLY-109 AND

RP VAL-131.

RC TISSUE=Brain;

RX MEDLINE=20053880; PubMed=10585853;

RA Sano S.-I., Ohnishi H., Kubota M.;

RT "Gene structure of mouse BIT/SHPS-1."

RL Biochem. J. 344:667-675(1999).

RN (5)

RP SEQUENCE FROM N.A.

RX MEDLINE=21638749; PubMed=11780052;

RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,

RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,

RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,

RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,

RA Buck D., Burrill W.D., Butler A.P., Gardner C., Carter N.P.,

RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,

RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,

RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,

RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,

RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E.,

RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,

RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,

RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,

RA Leharasalo M.H., Leverhwa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,

RA Marsh V.L., Martin S.L., McConachie L.J., McLeay K., Murray A.A.,

RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,

RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,

RA Phillimore B.J.C.T., Prichard S.R., Plumb R.W., Ramsay H.,

RA Rice C.M., Ross M.T., Scott C.E., Senra H.K., Showkhen R., Sims S.,

RA Skuse C.D., Smith M.L., Soderlund C., Stewart C.A., Sultson J.E.,

RA Swann R.M., Symmore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,

RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,

RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,

RA Wilmshurst L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,

RA Rogers J.;

RT "The DNA sequence and comparative analysis of human chromosome 20."

RL Nature 414:865-871(2001).

RN (6)

RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANTS LEU-44; THR-50;

RP THR-52; ARG-54; ALA-57; GLY-75; ASP-95; LEU-96; ASN-100; ARG-107;

RP GLY-109 AND VAL-131.

RC TISSUE=Brain, Kidney, and Skin;

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Scheer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Urdin T.B., Toshiyuki S., Cantucci P., Prange C.J.,

RA Rabe S.S., Loggiano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.U., Malek A.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W.,

RA Villalon D.K., Muzny K.D., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,

RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length

RT human and mouse cDNA sequences."

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN (7)

RP FUNCTION, AND INTERACTIONS WITH FYB; SCAP2 AND PTK2B.

RX MEDLINE=99401000; PubMed=10465959;

RA Tims J.F., Swanson K.D., Marie-Cardine A., Raab M., Rudd C.E.,

RA Schraven B., Neel B.G.;

RT "SHS-1 is a scaffold for assembling distinct adhesion-regulated

RT multi-protein complexes in macrophages."

RL Curr. Biol. 9:927-930(1999).

RN (8)

RP PHOSPHORYLATION BY JAK2, AND INTERACTIONS WITH PTPN11 AND JAK2.

RX MEDLINE=20428742; PubMed=10842184;

RA Stofega M.R., Argersinger L.S., Wang H., Ullrich A., Carter-Su C.;

RT "Negative regulation of growth hormone receptor/JAK2 signaling by

RT signal regulatory protein alpha."

RL J. Biol. Chem. 275:28222-28229(2000).

RN (9)

RP FUNCTION, AND INTERACTION WITH CD47.

RX MEDLINE=21400825; PubMed=11509594;

RA Latour S., Tanaka H., Demeure C., Mateo V., Rubio M., Brown E.J.,

RA Maliszewski C., Lindberg F.P., Oldenborg A., Ullrich A.,

RA Delapasse G., Sarfati M.;

RT "Bidirectional negative regulation of human T and dendritic cells by

RT CD47 and its cognate receptor signal-regulator protein-alpha:

RT down-regulation of IL-12 responsiveness and inhibition of dendritic

RT cell activation."

RL J. Immunol. 167:2547-2554(2001).

RN (10)

RP FUNCTION: Immunoglobulin-like cell surface receptor for CD47. Acts

as docking protein and induces translocation of PTPN6, PTPN11 and

other binding partners from the cytosol to the plasma membrane.

Supports adhesion of cerebellar neurons, neurite outgrowth and glial cell attachment. May play a key role in intracellular signaling during synaptogenesis and in synaptic function (By similarity). Involved in the negative regulation of receptor tyrosine kinase-coupled cellular responses induced by cell adhesion, growth factors or insulin. Mediates negative regulation of phagocytosis, mast cell activation and dendritic cell activation. CD47 binding prevents maturation of immature dendritic cells and inhibits cytokine production by mature dendritic cells.

-1- SUBUNIT: Binds PRPNI1 when tyrosine-phosphorylated, except in macrophages, where it primarily binds PRP6. Binds GSB2 in vitro. Binds FGR (by similarity). Binds JAK2 irrespective of its phosphorylation status and forms a stable complex. Binds SCAP1 and/or SCAP2. The resulting complex recruits Fyb. Binds PTK2B.

-1- SUBCELLULAR LOCATION: Type I membrane protein.

-1- ALTERNATIVE PRODUCTS:

Event-Alternative splicing; Named isoforms=3;

Name=1;

isoId=P78324-1; Sequence=Displayed;

Name=2;

isoId=P78324-2; Sequence=VSP_007030;

CC Note=No experimental confirmation available;

CC Name=3;

CC isoId=P78324-3; Sequence=VSP_007029;

CC Note=No experimental confirmation available;

CC -1- TISSUE SPECIFICITY: Ubiquitous. Highly expressed in brain. Detected on myeloid cells, but not T cells. Detected at lower levels in heart, placenta, lung, testis, ovary, colon, liver, small intestine, prostate, spleen, kidney, skeletal muscle and pancreas.

CC -1- PTM: N-glycosylated.

CC -1- PTM: Phosphorylated on tyrosine residues in response to stimulation with EGF, growth hormone, insulin and PDGF. Dephosphorylated by PTPN11.

CC -1- SIMILARITY: Contains 2 immunoglobulin-like C1-type domains.

CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: D86043; BAA12974.1; -

DR EMBL: Y10375; CAAT71403.1; -

DR EMBL: Y11047; CAAT71944.1; ALT_INT.

DR EMBL: AB023430; BAA87929.1; -

DR EMBL: AC004832; AAF19260.1; -

DR EMBL: AL034562; CAB38874.1; -

DR EMBL: AL049634; CAB46662.1; ALT_SEQ.

DR EMBL: AL117335; CAC12723.1; -

DR EMBL: BC026692; AAH26692.1; -

DR EMBL: BC033092; AAH33092.1; -

DR EMBL: BC038510; AAH38510.1; -

DR PIR: JCS287; JCS287.

DR Genew: HGNC:9662; PRPNS1.

DR MIM: 602461; -

DR GO: GO:0005886; C:plasma membrane; TMS.

Query Match 4.3%; Score 147.5; DB 1; Length 503;

Best Local Similarity 22.3%; Pred. No. 0.03;

Matches 88; Conservative 55; Mismatches 169; Indels 83; Gaps 17;

QY 259 KEVSKRYTQPKLQMGKPLPLHLPLPQALP-----QYAGSGLT-LTLEAKTGLHDEV 312

DB 32 BELQVITQPKSVSAAGSALIHCTVSLIPVGPQWFRGAPARELLYNGKEGF----- 87

QY 313 NLVVRATQLOKNLCEVWGPSTPLMLSLKLENKEAVSKREKPVWTLNPDAGWQCLL 372

DB 88 -----PRVTVSSTYKRENDPSISISN-----ITPADAGTYCVK 123

QY 373 SDSGQVLLSNIKVLPTWSPPEBKSCKDTH-TCPCPABELLGGSPVFLPPPKDQTL 431

DB 124 FRKG-----SPDTEFKSGAGTELIVRAKPSAPVVSPPARATP----- 161

QY 432 ISRPETVCVVVDVSHDEPVEKFMVYGVVHNAKTK--PREP-----QNSTRYVSVL 485

DB 162 -QHTVSTFCSHGSPRD--ITLKMFGNGELSPQTNVDPVGSVSYSIHSTAKVLT 218

QY 486 TVLHQDWLNGKEYCKVKNALPA-PIEKT--ISKAGQPREPOVYTLPPSRDELTKNOV 542

DB 219 EDVHSQVY-----CEFAHTLTQGLDPLRGTAISSETIRVPTLEV-TQGPYRAE---NQV 268

QY 543 SLTCLVGPVPSDIAVESNGQPENNYKTPPYLSDGSFFLYSKLTVDSRWQGNVF 602

DB 269 NVTCQVAKFPQRLQTLNLEGNVSRTEASTVENKDGTYNMWSMLLVNSAHRDVKL 328

QY 603 SCSTVME---ALNHYTKSLSLSPGLQDETCE 634

DB 329 TCQVHDGQPAVSKSHDK-VSAHPKQSGSNTAE 362

RESULT 89

SRBI_HUMAN STANDARD; PRT; 398 AA.

AC 000241; O8TB12; O9H1U5; Q9Y4V0;

DT 28-FEB-2003 (Rel. 41, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Signal-regulatory protein beta-1 precursor (SIRP-beta-1).

GN SIRPB1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RP [1]

RN SEQUENCE FROM N.A. (ISOFORM 1).

RC TISSUE=Placenta;

RX MEDLINE=97215901; PubMed=9062191;

RA Khartonenkov A., Chen Z., Suree I., Wang H., Schilling J., Ullrich A.;

RT "A family of proteins that inhibit signalling through tyrosine kinase receptors.";

RL Nature 386:181-186(1997).

RP [2]

RX SEQUENCE FROM N.A.

RE MEDLINE=21638749; PubMed=11780052;

RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R., Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L., Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M., Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J., Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M., Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R., Coulson A., Coville G.J., Deadman R., Dham P.D., Dunn M., Ellington A.G., Frankland J.A., Fraser A., French L., Garner P., Griffith D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E., Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J., Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D., Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Laylor S., Levanon D.H., Levanon M.H., Levanon M.A., Lloyd C., Lloyd D.M., Lovell A.J.D., Marsh V.L., Martin S.L., McCormack L.J., McElay K., McMurtry A.A., Milne S.A., Mistry A., Moore M.J.F., Mullikin J.C., Nickerson T., Oliver K.A., Parker A., Patel R., Pearce T.A.V., Peck A.I., Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsey H., Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S., Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E., Swann R.M., Symcote N., Taylor R., Tee L., Thomas D.W., Thorpe A., Trexan A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Whitehead S.L., Whitteker P., Willey D.L., Williams L., Williams S.A., Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S., Rogers J.;

RT "The DNA sequence and comparative analysis of human chromosome 20.";

RL Nature 414:865-871(2001).

[3]
SEQUENCE FROM N.A. (ISOFORM 2).
TISSUE=Skin;
MEDLINE=22388257; PubMed=12477932;
Strauberg R. L., Feingold E. A., Grouse L. H., Derge J. G.,
Klausner R. D., Collins F. S., Wagner L., Shemen C. M., Schuler G. D.,
Altschul S. F., Zeeberg B., Buetow K. H., Schaefer C. F., Blatt N. K.,
Hopkins R. F., Jordan H., Moore T., Max S. I., Wang J., Hsieh P.,
Diatchenko L., Marusina K., Farmer A. A., Rubin G. M., Hong L.,
Stapleton M., Soares M. B., Bonaldi M. F., Casavant T. L., Scheetz T. E.,
Brownstein M. J., Ueda T. B., Toshiyuki S., Carninci P., Prange C.,
Raha S. S., Loguettano N. A., Peters G. J., Abramson R. D., Mullaly S. J.,
Bosnak S. A., McEwen P. J., McKernan K. J., Malek J. A., Gnatratte P. H.,
Richards S., Worley K. C., Hale S., Garcia A. M., Gay L. J., Hulyk S. W.,
Villalon D. K., Muzny D. M., Sodergren E. J., Lu X., Gibbs R. A.,
Fahley J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A. C., Shevchenko Y., Bouffard G. G.,
Blakeley R. W., Touchman J. W., Green E. D., Dickson M. C.,
Rodriguez A. C., Grimwood J., Schmutz J., Myers R. M.,
Butterfield Y. S. N., Krzywinski M. I., Skalska U., Smalins D. E.,
Schnerich A., Schain J. E., Jones S. J. M., Marra M. A.;
"Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[4]
INTERACTION WITH TYROBP AND SYK.
PubMed=11694422;
RA Tomassello E., Cant C., Buehring H.-J., Vely F., Andre P., Seifert M.,
RA Ullrich A., Vvieler E.;
RT "Association of signal-regulatory proteins beta with KARPAP/DAP-12.";
Eur. J. Immunol. 30:2147-2156(2000).
[5]
INTERACTION WITH TYROBP, N-GLYCOSYLATION, AND TISSUE SPECIFICITY.
PubMed=20072721; PubMed=10604985;
RA Dietrich J., Cella M., Seifert M., Buehring H.-J., Colonna M.;
RT "Signal-regulatory protein beta 1 is a DAP12-associated activating
receptor expressed in myeloid cells.";
J. Immunol. 164:9-12(2000).
-1- FUNCTION: Immunoglobulin-like cell surface receptor involved in
the negative regulation of receptor tyrosine kinase-coupled
signaling processes. Participates also in the recruitment of
tyrosine kinase SYK.
CC -1- SUBUNIT: Interacts with TYROBP. This interaction results in the
recruitment of SYK.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=1;
CC -1- IsoId=000241-1; Sequence=Displayed;
Name=2;
CC IsoId=000241-2; Sequence=VSP 007026;
CC Note=No experimental confirmation available;
CC -1- TISSUE SPECIFICITY: Detected in monocytes and dendritic cells.
CC -1- PTM: N-glycosylated.
CC -1- SIMILARITY: Contains 2 immunoglobulin-like C1-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch).
CC EMBL, Y10376; CAA71404.1; -
CC EMBL, AL049634; CAB4661.2; -
CC EMBL, AL138804; CAC17540.1; -
CC EMBL, BC025286; AAH25286.1; -
CC Genew; HGNC:15928; SIRPB1.
CC MIM, 603889; -
CC GO, GO:0005887; C:Integral to plasma membrane; TAS.
CC GO, GO:0007166; P:cell surface receptor linked signal transdu. .; TAS.

DR	InterPro; IPRO007110; Ig-like.
DR	InterPro; IPRO03597; Ig-cl.
DR	Pfam; PF00047; Ig. 3.
DR	SMART; SMO0407; IgcI. 2.
DR	SMART; PS00835; IG LIKE; 3.
KW	Repeat; Signal; Transmembrane; Immunoglobulin domain; Glycoprotein;
FT	Alternative splicing.
FT	SIGNAL. 26
FT	CHAIN. 27
FT	DOMAIN. 27
FT	DOMAIN. 27
FT	TRANSMEM. 372
FT	DOMAIN. 393
FT	DOMAIN. 27
FT	DOMAIN. 147
FT	DOMAIN. 253
FT	DISULFID. 54
FT	DISULFID. 169
FT	DISULFID. 272
FT	CARBOHYD. 244
FT	CARBOHYD. 269
FT	CARBOHYD. 291
FT	VARSPLIC. 145
FT	VARSPLIC. 361
FT	CONFLICT. 53
FT	CONFLICT. 102
FT	CONFLICT. 229
FT	CONFLICT. 363
FT	CONFLICT. 363
SO	SEQUENCE. 398 AA; 43255 MW; A2MA08FE8E2BC52B CXC64;
Qy	Query Match. 4.3%; Score 147; DB 1; Length 398;
Db	Best Local Similarity. 24.7%; Pred. No. 0.024;
Matches	57; Conservative. 41; Mismatches. 93; Indels. 40; Gaps. 10;
Qy	409 PABELLGGPSYFLFPKKPKDTIMSRPEVCVVVDVSH--EDPEKFNMYDVGEVHNA 466
Db	147 PSAPVSPAV-----RATPEHTVVSFCSEHGSSPRDITLMKFNGLSDF 193
Qy	467 KTK--PRE-----QYNSTRVSVTLVLHDMLNGKEKYCKVSKALPA-PIEKT-ISK 517
Db	194 QTNVPADGDSVSYSHSTARVTLRGDVHSQVI-----CEMAHITLGDDPLRGTAISE 247
Qy	518 AKGPREFEQVYTLPPSRBELTKNQVSLTCLYGFPSPIAIVBSNCGEPENNYKTPPVL 577
Db	248 AIRVPEPTLEV-TQQPMRAE---NOANVTCCGVSNFPRGLQTWLGNVSTRRTASTLSIE 303
Qy	578 DSDSFFLYSLTLDVKSRMGQGAFSCGYMEHALHNHTOKSLSLSPGLQL 628
Db	304 NKDGTYMMWSLVLTNTCAHRDDVVLTCTVEHDG-----QDAVSKSYALEI 348
RESULT 90	
ID	SNEZ_HUMAN STANDARD. PRT; 6885 AA.
AC	O8MXH0; O8NI13; O8NF49; O8TER7; O8MW3; O8MW4; O8MW5; O8MXH1;
AC	O8NU05; O9UFQ4; O9YZL4; O9Y4R1;
DT	10-OCT-2003 (Rel. 42, Created)
DT	10-OCT-2003 (Rel. 42, Last sequence update)
DT	10-OCT-2003 (Rel. 42, Last annotation update)
DE	Nesprin 2 (Nuclear envelope spectrin repeat protein 2) (Syne-2)
DE	(synaptic nuclear envelope protein 2) (Nucleus and actin connecting
DE	element protein) (NUANCE protein).
OS	Glycine or NUA OR KIA1011.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Euthera; Primates; Catarrhini; Homidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A. (ISOFORMS 1 AND 9), FUNCTION, CHARACTERIZATION, AND
RP	INTERACTION WITH F-ACTIN.
RX	MEDLINE=22113122; PubMed=12118075;
RA	Zhen Y.-Y., Libotte T., Munck M., Noegel A.A., Korenbaum E.;
RT	"NUANCE, a giant protein connecting the nucleus and actin
RT	cytoskeleton."/;

RL J. Cell Sci. 115:3207-3222(2002).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 4, 5 AND 7).
 RX MEDLINE=21652858; PubMed=11792814;
 RA Zhang Q., Skepper J.N., Yang F., Davies J.D., Heyl L., Roberts R.G.,
 RA Weissberg P.L., Ellis J.A., Shanahan C.M.;
 RT "Neaprin: a novel family of spectrin-repeat-containing proteins that
 RT localise to the nuclear membrane in multiple tissues.";
 RL J. Cell Sci. 114:4485-4498(2001).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1)
 RX MEDLINE=22296983; PubMed=12408964;
 RA Zhang Q., Ragnauth C., Greener M.J., Shanahan C.M., Roberts R.G.;
 RT "The neaprinins are giant actin-binding proteins, orthologous to
 RT Drosophila melanogaster muscle protein MSP-100.";
 RL Genomics 80:473-481(2002).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 6).
 RX TISSUE=Testis;
 RA Poustka A., Klein M., Mewes H.-W., Gaassenhuber J., Wiemann S.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RX PubMed=12508121;
 RA Heilig R., Eckenberg R., Petit J.-L., Fonknechten N., Da Silva C.,
 RA Catellon L., Levy M., Barde V., De Bernardis V., Ureta-Vidal A.,
 RA Pelletier E., Vico V., Anthouard V., Rowen L., Madan A., Qin S.,
 RA Sun H., Du H., Pepin K., Artiguenave F., Robert C., Cruaud C.,
 RA Brule T., Jallouin O., Friedlander L., Samson G., Brocquier P.,
 RA Cure S., Segurens B., Aniere F., Samain S., Crespeau H., Abbasi N.,
 RA Alach N., Boescu D., Dickhoff R., Dots M., Dubois I., Friedmann C.,
 RA Gouvenoux M., James R., Madan A., Mairey-Bertrida B., Mangot S.,
 RA Martine N., Menard M., Oztas S., Ratcliffe A., Shaffer T., Traek B.,
 RA Vachier B., Bellemere C., Belser C., Beaumont-Cornet M.,
 RA Batrol-Mavel D., Bourdard M., Brier-Silla S., Combette S.,
 RA Dufosse-Laurent V., Ferron C., Lechaplais C., Lousse C., Musset D.,
 RA Magdelan G., Patrau E., Petit E., Sirvain-Trukniewicz P., Trybou A.,
 RA Vega-Carriy N., Bataille E., Bluet E., Bordelais I., Dubois M.,
 RA Dumont C., Guerin T., Haffray S., Hamadi R., Munga J., Pellouin V.,
 RA Robert D., Munderle B., Gauguet G., Roy A., Sainte-Marthe L.,
 RA Verdier J., Verdier-Discala C., Hillier L.W., Fulton L., McPherson J.,
 RA Marcuda F., Wilson R., Scarpelli C., Gyapay G., Wincker P., Saurin W.,
 RA Queller F., Waterston R., Hood L., Weissenbach J.;
 RT "The DNA sequence and analysis of human chromosome 14.";
 RL Nature 421:601-607(2003).
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM 8).
 RX TISSUE=Brain;
 RA MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altschul S.F., Zeebner B., Buetow K.H., Scheffer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.W., Hong L.,
 RA Steglein M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carroll P., Prange C.,
 RA Rana S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Boeck S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Halyk S.W.,
 RA Villalón D.K., Muzny K.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield V.S.N., Krzywinski M.T., Skalska U., Smallus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.W., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [7]
 RP SEQUENCE FROM N.A. (ISOFORM 3), AND SEQUENCE OF 1-956 AND 5133-6885
 RP FROM N.A.
 RC TISSUE=Spleen, and Tongue;

RA Jikuya H., Takano J., Nomura N., Kikuno R., Nagase T., Ohara O.,
 RA Nimomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
 RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
 RA Katsura N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
 RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
 RA Suuki Y., Sugano S., Nagahari K., Masuo Y., Nagai K., Isogai T.;
 RT "NEO human cDNA sequencing project.";
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RP SEQUENCE OF 5709-6885 FROM N.A. (ISOFORM 2).
 RX TISSUE=Brain;
 RA MEDLINE=99246063; PubMed=10231032;
 RC TISSUE=Brain;
 RX MEDLINE=99246063; PubMed=10231032;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.,
 RT "Prediction of the coding sequences of unidentified human genes. XIII.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 6:63-70(1999).
 RN [9]
 RP REVISIONS.
 RC TISSUE=Brain;
 RX MEDLINE=22158633; PubMed=12168954;
 RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
 RT "Construction of expression-ready cDNA clones for KIAA genes: manual
 RT curation of 330 KIAA cDNA clones.";
 RL DNA Res. 9:99-106(2002).
 RN [10]
 RP SEQUENCE OF 5754-6885 FROM N.A.
 RX MEDLINE=21154917; PubMed=11230166;
 RA Wiemann S., Weil B., Wellenreuther R., Gaassenhuber J., Glaesl S.,
 RA Ansgore W., Boecker M., Bioecker H., Bauesachs S., Blum H.,
 RA Laber J., Duesenhoelt A., Beyer A., Koehner K., Strack N.,
 RA Mewes H.-W., Oetemaelder B., Obermaier B., Tampe J., Heubner D.,
 RA Wandus R., Korn B., Klein M., Poustka A.;
 RT "Towards a catalog of human genes and proteins: sequencing and
 RT analysis of 500 novel complete protein coding human cDNAs.";
 RL Genome Res. 11:422-435(2001).
 RN [11]
 RP -1- FUNCTION: Involved in the maintenance of nuclear organization and
 RP structural integrity. Probable anchoring protein which tethers the
 RP nucleus to the cytoskeleton. Connects nuclei to the cytoskeleton
 RP by interacting with the nuclear envelope and with F-actin in the
 RP cytoplasm.
 RN [12]
 RP -1- SUBUNIT: Interacts with F-actin via its N-terminal domain.
 RP -1- SUBCELLULAR LOCATION: Type IV membrane protein (Potential). The
 RP largest part of the protein is cytoplasmic, while its C-terminal
 RP part is associated with the nuclear envelope, most probably the
 RP outer nuclear membrane. Remains associated with the nuclear
 RP envelope during its breakdown in mitotic cells.
 RN [13]
 RP -1- ALTERNATIVE PRODUCTS:
 RP Event=Alternative splicing; Named isoforms=9;
 RN Name=1;
 CC IsoId=08WXH0-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=08WXH0-2; Sequence=VSP_007164, VSP_007166;
 CC Note=No experimental confirmation available;
 CC Name=3;
 CC IsoId=08WXH0-3; Sequence=VSP_007155;
 CC Note=Produced by exon skipping that results in a frameshift. No
 CC experimental confirmation available;
 CC Name=4; Synonyms=Beta;
 CC IsoId=08WXH0-4; Sequence=VSP_007156;
 CC Name=5; Synonyms=Alpha;
 CC IsoId=08WXH0-5; Sequence=VSP_007157, VSP_007164, VSP_007165;
 CC Name=6;
 CC IsoId=08WXH0-6; Sequence=VSP_007158, VSP_007165, VSP_007166;
 CC Note=No experimental confirmation available;
 CC Name=7; Synonyms=Gamma;
 CC IsoId=08WXH0-7; Sequence=VSP_007154, VSP_007163;
 CC Name=8;
 CC IsoId=08WXH0-8; Sequence=VSP_007161, VSP_007162;

DT 10-OCT-2003 (Rel. 42, last sequence update)
DE 10-OCT-2003 (Rel. 42, last annotation update)
DE Signal-regulatory protein beta-2 precursor (SIRP-beta-2) (SIRP-b2).
CN SIRPB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=21036165; PubMed=1185750;
RA Ichigotani Y., Matsuda S., Machida K., Oshima K., Iwamoto T.,
RA Yamaki K., Hayakawa T., Hamaguchi M.;
RT "Molecular cloning of a novel human gene (SIRP-B2) which encodes a new
RT member of the SIRP/SHPS-1 protein family.";
RL J. Hum. Genet. 45:378-382(2000).
[2]
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggaley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Frazer A., French L., Garner P.,
RA Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Levanthal M.H., Leverhwa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McElay K., McKerrry A.A.,
RA Milne K., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prichaltingam S.R., Plumb R.W., Ramsey H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkhen R., Sims S.,
RA Skuse C.D., Smith M.L., Soderlund C., Steward C.A., Sulterson J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Thorpe A.,
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
[3]
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932;
RA Struhsberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M.J., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ussid T.B., Toshiyuki S., Carinini P., Prange S.J.,
RA Raba S.S., Loguclano N.A., Peters G.J., Abrahamson R.D., Mullaly S.J.,
RA Boeck S.A., McEwan P.J., McKernan K.J., Malek J.A., Gamarate P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hilyk S.W.,
RA Villalón D.K., Wozny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Faley J., Helton E., Kerteman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Scheraga A., Schin J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- FUNCTION: Probable immunoglobulin-like cell surface receptor.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=Q9PIW8-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9PIW8-2; Sequence=VSP_007027;
CC Note=No experimental confirmation available;
CC Name=3;
CC IsoId=Q9PIW8-3; Sequence=VSP_007028;
CC Note=No experimental confirmation available;
CC -1- TISSUE SPECIFICITY: Detected in liver, and at very low levels in
CC brain, heart, lung, pancreas, kidney, placenta and skeletal
CC muscle.
CC -1- SIMILARITY: Contains 2 immunoglobulin-like V-type domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; AB042624; BAA95692.1; -;
CC EMBL; AL138804; CAC00474.1; -;
CC EMBL; BC020629; AAH0629.1; ALT_INT.
CC Genew; HGNC:15757; SIRPB2.
CC MIM; 605466; -;
DR GO; GO:0007267; P:cell-cell signaling; TAS.
DR GO; GO:0007242; P:intracellular signaling cascade; TAS.
DR GO; GO:0008285; P:negative regulation of cell proliferation; TAS.
DR InterPro; IPR003597; IG-like.
DR InterPro; IPR003597; IG-like.
DR InterPro; IPR003597; IG-like.
DR InterPro; IPR003596; IG-V.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00407; IG1; 2.
DR SMART; SM00406; IG1; 1.
DR PROSITE; PS00835; IG-like; 3.
DR PROSITE; PS00290; IG-MHC; FALSE NEG.
DR Repeat; Signal; Transmembrane; Immunoglobulin domain; Glycoprotein;
KM Alternative splicing.
FT SIGNAL 1 28
FT CHAIN 29 387
FT DOMAIN 29 360
FT TRANSMEM 361 383
FT DOMAIN 384 387
FT DOMAIN 29 137
FT DOMAIN 146 245
FT DOMAIN 252 340
FT DISULFID 53 119
FT DISULFID 168 226
FT DISULFID 271 329
FT CARBOHYD 243 243
FT CARBOHYD 268 268
FT CARBOHYD 309 309
FT CARBOHYD 317 317
FT VARSPLIC 1 33
FT VARSPLIC 144 360
FT VARSPLIC 263 263
FT CONFLICT 286 286
FT SEQUENCE 387 AA; 42495 MW; F7F20C9F86B0E64B CRC64;
Query Match 4.3%; Score 146; DB 1; Length 387;
Best Local Similarity 24.8%; Pred. No. 0.027;
Matches 57; Conservative 37; Mismatches 114; Indels 22; Gaps 8;
Db 412 ELARGPVS-FLPPKPKDTLMIS----RTPEVTCVVDVSH--BDPEVKRMVYDVCVEVH 464
131 EFKSGPTETALGAKPAPVILGPAARTPEHTVSTFCESHGSPFRDITLWFKNGNELS 190

```

QY 465 NAKTKREBOYNSTYRVVSLTVLHODWLNKGEYKCKNSKALPA-PIEKT--ISKAKQ 521
DB 191 DQOTNDVPTPGQSVANSIRSTARVLDPMVRSQVCEVAVHTLQGDPLGLTANLSEAIRV 250
QY 522 PREPOVYTLPPSRDELTKNOVSLTCLVKGFPSPDIWESNGOPENNYKTPPVLDSPG 581
DB 251 PPTLEV-TQGPKR---AGNOVNTVQVRKFFYQSLQTLWLENGNVCGQRETAATLTENKDG 306
QY 582 SFPLYSKLTVDKSRMOQGNVFCSCVMHKLHNHYTKSLSPGLQDFT 631
DB 307 TYNMTSMFLVNIISDQRDVVLTCOVGHDC-----QLAVSKRLALEVT 348

RESULT 93
LAC_CHICK
ID LAC_CHICK STANDARD; PRT; 103 AA.
AC P20763;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DS Ig lambda chain C region.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN (1)
RX MEDLINE=87218480; PubMed=3107981;
RX Parvael R., Ziv E., Lenterer F., Tel-Or S., Bursstein Y., Schechter I.;
RT "Analyses of chicken immunoglobulin light chain cDNA clones indicate
RT a few germline V lambda genes and allotypes of the C lambda locus.";
RL EMBO J. 6:37-102(1987).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; X04768; CA28461.1; -.
CC DR PIR; B26167; B26167.
CC DR HSSP; P01842; 7FAB.
CC DR InterPro; IPR007110; IG-1like.
CC DR InterPro; IPR003597; IG_c1.
CC DR InterPro; IPR003006; IG_MHC.
CC DR Pfam; PF00047; Ig_1.
CC DR SMART; SM00407; IGc1; 1.
CC DR PROSITE; PS00835; IG_LIKE; 1.
CC DR PROSITE; PS00290; IG_MHC; 1.
CC KM Immunoglobulin domain; Immunoglobulin C region.
CC FT NON_TER 1 1
CC FT DOMAIN 6 99 IG-LIKE.
CC FT DISULFID 28 85
CC FT DISULFID 103 103 INTERCHAIN (WITH HEAVY CHAIN).
CC FT VARIANT 90 90 N -> D.
CC SO SEQUENCE 103 AA; 11361 MW; 77BF341B511B91B2 CRC64;

Query Match 4.3%; Score 145.5; DB 1; Length 103;
Best Local Similarity 32.2%; Pred. No. 0.005;
Matches 29; Conservative 19; Mismatches 39; Indels 3; Gaps 3;

QY 521 OR-EPQVYTLPPSRDELTK-NOVSLTCLVKGFPSPDIWESNGOPENNYKTPPVLD 578
DB 1 QPVAATITLFPSPSKELNATKATLVCLINDYPSVTVVDVDSITSG-ETTAPOQO 59
QY 579 SDGSFPLYSKLTVDKSRMOQGNVFCSCVMH 608
DB 60 SNSQVYASSYLSASDMSHETTCRYTH 89

```

```

RESULT 94
KACB_RAT
ID KACB_RAT STANDARD; PRT; 106 AA.
AC P01835;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DS Ig kappa chain C region, B allele.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN (1)
RX MEDLINE=75212238; PubMed=807630;
RX Starace V., Querinjean P.;
RT "The primary structure of a rat kappa Bence Jones protein:
RT phylogenetic relationships of V- and C-region genes.";
RL J. Immunol. 115:59-62(1975).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC DR HSSP; P01842; 2MCG.
CC DR InterPro; IPR007110; Ig-1like.
CC DR InterPro; IPR003597; IG_c1.
CC DR Pfam; PF00047; Ig_1.
CC DR SMART; SM00407; IGc1; 1.
CC DR PROSITE; PS00835; IG_LIKE; 1.
CC DR PROSITE; PS00290; IG_MHC; 1.
CC KM Immunoglobulin domain; Immunoglobulin C region.
CC FT NON_TER 1 1
CC FT DOMAIN 5 102 IG-LIKE.
CC FT DISULFID 26 86
CC FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
CC FT CONFLICT 2 2 D -> N (IN REF. 2).
CC FT CONFLICT 30 30 N -> K (IN REF. 2).
CC FT CONFLICT 48 48 MISSING (IN REF. 2).
CC FT CONFLICT 79 79 E -> Q (IN REF. 2).
CC FT CONFLICT 87 87 E -> Q (IN REF. 2).
CC FT CONFLICT 98 98 V -> VM (IN REF. 2).
CC FT CONFLICT 100 100 S -> N (IN REF. 2).
CC SO SEQUENCE 106 AA; 11601 MW; 4CFA7CA820D1CA36 CRC64;

Query Match 4.2%; Score 145; DB 1; Length 106;
Best Local Similarity 31.5%; Pred. No. 0.0056;
Matches 29; Conservative 22; Mismatches 23; Indels 12; Gaps 2;

QY 525 POVYTLPPSRDELTKNOVSLTCLVKGFPSPDIWESNGOPENNYKTPPVLDSPG 579
DB 5 PTVSIFPSTPQLATGASVVCCLMNNFYPRDISYKWKIDGTERRD-----CVLDSVTDQD 59
QY 580 --DGSFPLYSKLTVDKSRMOQGNVFCSCVMH 609
DB 60 SKDSTYSMSSTLSLKADYESHNLVTEVAK 91

RESULT 95
C166_CARAU
ID C166_CARAU STANDARD; PRT; 555 AA.
AC Q90304;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

```


RC TISSUE=Brain;
 RX MEDLINE=92120663; PubMed=1769655;
 RA Hlavayn M.L., Lemmon V.;
 RT "Molecular structure and functional testing of human LICAM: an
 RT interspecies comparison.";
 RL Genomics 11:416-423(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92329299; PubMed=1627459;
 RA Reid R.A., Hemperly J.J.;
 RT "Variants of human L1 cell adhesion molecule arise through alternate
 RT splicing of RNA.";
 RL J. Mol. Neurosci. 3:127-135(1992).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Rosenthal A., Couelle O., Dreacher B.;
 RL Submitted (Apr-1994) to the EMBL/Genbank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97432815; PubMed=9286695;
 RA Brenner V., Nyakatura G., Rosenthal A., Platzner M.;
 RT "Genomic organization of two novel genes on human Xq28: compact head
 RT to head arrangement of IDH gamma and TRAP delta is conserved in rat
 RT and mouse.";
 RL Genomics 44:8-14(1997).
 RN [6]
 RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
 RX MEDLINE=98147998; PubMed=9479034;
 RA Couelle O., Nyakatura G., Tauden S., Elgar G., Brenner S.,
 RA Platzner M., Dreacher B., Jouet M., Kenwick S., Rosenthal A.;
 RT "The neural cell adhesion molecule L1: genomic organisation and
 RT differential splicing is conserved between man and the pufferfish
 RT Fugu.";
 RL Gene 208:7-15(1998).
 RN [7]
 RP SEQUENCE OF 20-36.
 RX MEDLINE=88298876; PubMed=3136168;
 RA Wolff J.M., Frank R., Mujoo K., Spiro R.C., Relafeld R.A.,
 RA Rathjen F.G.;
 RT "A human brain glycoprotein related to the mouse cell adhesion
 RT molecule L1.";
 RL J. Biol. Chem. 263:11943-11947(1988).
 RN [8]
 RP SEQUENCE OF 332-371 FROM N.A.
 RX MEDLINE=90353957; PubMed=2387585;
 RA Djafarzadeh M., Matzel M.-G., Nguyen C., Roux D., Demergeot J.,
 RA Denzot F., Moos M., Schachner M., Goridis C., Jordan B.R.;
 RT "The gene encoding L1, a neural adhesion molecule of the
 RT immunoglobulin family, is located on the X chromosome in mouse and
 RT man.";
 RL Genomics 7:587-593(1990).
 RN [9]
 RP SEQUENCE OF 353-1176 FROM N.A.
 RX TISSUE=Fetal brain;
 RX MEDLINE=92020233; PubMed=1923824;
 RA Rosenthal A., Mackinnon R.N., Jones D.S.C.;
 RT "PCR walking from microdissection clone M54 identifies three exons
 RT from the human gene for the neural cell adhesion molecule L1
 RT (CAN-L1).";
 RL Nucleic Acids Res. 19:5395-5401(1991).
 RN [10]
 RP SEQUENCE OF 809-1257 FROM N.A.
 RX TISSUE=Pancreas;
 RX MEDLINE=92288257; PubMed=12477932;
 RA Klausner R.L., Feilngold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhac N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.D., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalske U., Smalins D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [11]
 RP SEQUENCE OF 1030-1257 FROM N.A.
 RX MEDLINE=91132183; PubMed=1993895;
 RA Harper J.R., Prince J.T., Healy P.A., Stuart J.K., Nauman S.J.,
 RA Stalcup W.B.;
 RT "Isolation and sequence of partial cDNA clones of human L1: homology
 RT of human and rodent L1 in the cytoplasmic region.";
 RL J. Neurochem. 56:797-804(1991).
 RN [12]
 RP VARIANT HSAS TYR-264.
 RX MEDLINE=94004956; PubMed=8401576;
 RA Jouet M., Rosenthal A., Macfarlane J., Kenwick S., Donnai D.;
 RT "A missense mutation confirms the L1 defect in X-linked hydrocephalus
 RT (HSAS).";
 RL Nat. Genet. 4:331-331(1993).
 RN [13]
 RP VARIANT HSAS/MASA LEU-1194.
 RX MEDLINE=95187172; PubMed=7881431;
 RA Franssen E., Schrander-Stumpel C., Vits L., Coucke P., van Camp G.,
 RA Willems P.J.;
 RT "X-linked hydrocephalus and MASA syndrome present in one family are
 RT due to a single missense mutation in exon 28 of the LICAM gene.";
 RL Hum. Mol. Genet. 3:2255-2256(1994).
 RN [14]
 RP VARIANTS HSAS GLN-184 AND ARG-452, AND VARIANT MASA GLN-210.
 RX MEDLINE=95004608; PubMed=7920659;
 RA Jouet M., Rosenthal A., Armstrong G., Macfarlane J., Stevenson R.,
 RA Paterson J., Metzenberg A., Ionaescu V., Temple K., Kenrick S.;
 RT "X-linked spastic paraplegia (Sppl), MASA syndrome and X-linked
 RT hydrocephalus result from mutations in the L1 gene.";
 RL Nat. Genet. 7:402-407(1994).
 RN [15]
 RP VARIANTS MASA GLN-210 AND ASN-598.
 RX MEDLINE=95004609; PubMed=7920660;
 RA Vits L., van Camp G., Coucke P., Franssen E., de Bouille K.,
 RA Vitters E., Korn B., Poustcka A., Wilson G., Schrander-Stumpel C.,
 RA Winter R.M., Schwartz C., Willems P.J.;
 RT "MASA syndrome is due to mutations in the neural cell adhesion gene
 RT LICAM.";
 RL Nat. Genet. 7:408-413(1994).
 RN [16]
 RP VARIANTS HSAS/MASA SER-9; SER-121; LYS-309; PHE-768; LEU-941 AND
 RP CYS-1070.
 RX MEDLINE=95282776; PubMed=7762552;
 RA Jouet M., Moncla A., Paterson J., McKeown C., Fryer A., Carpenter N.,
 RA Holmberg E., Madelin C., Kenrick S.;
 RT "New domains of neural cell-adhesion molecule L1 implicated in
 RT X-linked hydrocephalus and MASA syndrome.";
 RL Am. J. Hum. Genet. 56:1304-1314(1995).
 RN [17]
 RP VARIANTS HSAS/MASA GLN-184, GLN-210, TYR-264; ARG-452; ASN-598 AND
 RP LEU-1194.
 RX MEDLINE=96153146; PubMed=8556302;
 RA Franssen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;
 RT "CRASH syndrome: clinical spectrum of corpus callosum hypoplasia,
 RT retardation, adducted thumbs, spastic paraparesis and hydrocephalus
 RT due to mutations in one single gene, L1.";
 RL Eur. J. Hum. Genet. 3:273-284(1995).
 RN [18]
 RP ERRORTW.
 RA Franssen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;
 RL Eur. J. Hum. Genet. 4:126-126(1996).

```

RN [19]
RP VARIANTS HSAS/MASA/SPG1 SER-179 AND ARG-370.
RX MEDLINE=96057511; PubMed=7562969;
RA Ruiz J.C., Cuppens H., Legius E., Fryns J.-P., Glover T., Marynen P.,
CA Caestejan J.-J.,
RT "Mutations in L1-CAM in two families with X linked complicated
RL epastic paraplegia, MASA syndrome, and HSAS."
RL J. Med. Genet. 32:549-552(1995).
[20]
RN VARIANTS HSAS CVS-194 AND LEU-240.
RX MEDLINE=97083370; PubMed=8929944;
RA Gu S.-M., Orth U., Veske A., Enders H., Klüender K., Schloesser M.,
RA Engel W., Schwinger E., Gal A.;
RT "Five novel mutations in the L1CAM gene in families with X linked
RT hydrocephalus."
RL J. Med. Genet. 33:103-106(1996).
[21]
RN VARIANTS HSAS GLN-184; 439-VAL--THR-443 DEL; CVS-784 AND
RP 936-LEU--LEU-948 DEL.
RX MEDLINE=97338664; PubMed=9195224;
RA Macfarlane J.R., Du J.-S., Peyre M.E., Ramezan S., Donnai D.,
RA Charlton R., Garrett C., Tolmie J., Yates J.R.W., Berry C., Goudie D.,
RA Monica A., Hunt P., Hodgson S., Jout M., Kenwick S.;
RT "Nine novel L1 CAM mutations in families with X-linked
RT hydrocephalus."
RL Hum. Mutat. 9:512-518(1997).
[22]
RN VARIANTS HSAS/MASA ASP-691; ARG-698 AND PRO-935.
RX MEDLINE=98180721; PubMed=9521424;
RA Du Y.-Z., Srivastava A.K., Schwartz C.E.;
RT "Multiple exon screening using restriction endonuclease
RT fingerprinting (REF) detection of six novel mutations in the L1 cell
RT adhesion molecule (L1CAM) gene."

```

Query Match 4.2%; Score 144.5; DB 1; Length 1257;

Best local Similarity 18.6%; Pred. No. 0.16;

Matches 129; Conservative 83; Mismatches 231; Indels 251; Gaps 32;

```

QY 8 RHLLLV-----LQALLPA--ATQGNKVLGKKGDTVELTCTASQKSIQFMKNSNQKI 61
DB 409 RHGLLAAAYIVVQLPAKILTDNQTYMVGSTAYILCAFGAPVSVOMLDEDTGV 468
QY 62 LGNGSFLTKGPSKLNDRADSRSLMDGPNFLIINKIKEDSPYICEVEDQEEVQL 121
DB 469 LQDERFF-----PYANGTLGRDQANDTGRYFCLANDQNNVTIM 509
QY 122 VFGLTANDTHLQ-----GSLTLTLES--PPSSSVQCRSRGNIQ----- 164
DB 510 A-NLKVKDATQITQPRSTIEKKSGRVFTQCAFSPSLQPSITWRGD-GRDLQELGDS 567
QY 165 -----GKTLVSQLELDGSGTWCTVLQNKQVFKIDIVLAF-----OKASSIYKK 214
DB 568 KYFIEDGR-LVTHSLDYSDQGNYSCTVASTELDVESRQQLLVGSPGVPRLVSDHL 626
QY 215 EGEQVESFPLA-----FTVEKLTSGELMWAQERASSKSNITFDLNKKEVS 263
DB 627 TQSGRVASMSPAEDHNAPIEKYDIEFEDKEMAPEKWSLKGPNQGTSTLKLSPYHYT 686
QY 264 KRYVQDPKLGKGLPLHLTL--POLAQ-----YASGNITLALAKTGLHBEVNLV 316
DB 687 FRVAVINKYGPGEPSVSETVVTPEAAPEKPNVDKVGNET-----TMMVI 733
QY 317 MRATQLQKNTLCEWGP-----TSPKMLSLKLENKAKVSKREKPVVPLNPAQMWO- 369
DB 734 -----TKKPLRMMDNAPQVQYKQ-----WRQGTGRGPRQE 765
QY 370 CLLSDSGVLLSNIKVLPTWSTPYEPKSCDHTCPCAPPELLGSPVFLFPKPKXT 429
DB 766 QIVSD--PFLVVSNTSTFVPEIKQAVNSQK-----GPE----- 799
QY 430 LMISTPEVTCVVVDVSHED-----PEYKFMVYVUGVEVHN-----AKTKR 471
DB 800 -----PQVT--IGYSGEDYPOALPE-----LEGIIINSSAVLVKMRPVDLAOVGH 844

```

```

QY 472 EEOYNSTRVSVLTLVHODPLNGKEYK-----CKVSNKALPA----- 509
DB 845 LRQNVNLY-----WREGSKRKHSHKHODVVVPANTTSVILSLRPSYSY 892
QY 510 -----PIEKTISKAGQPREPOVYTL-----PSRDELTKNOVS 543
DB 893 LEVOAFNGRSGPASEFTFSTPEGVGHPPEALHLECOSNTSLLRMQPP-----LSHNGV- 947
QY 544 LTCLVKGFPSPDIAVENESNGOPENNYKTPPVYL 577
DB 948 LTGYVLSYHPLD-----EGCKQGLSFNLRDPEL 975

```

RESULT 97

```

CAML_MOUSE
ID CAML_MOUSE STANDARD; PRT; 1260 AA.
AC P11627;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Neural cell adhesion molecule L1 precursor (N-CAM L1).
GN L1CAM OR CAML1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=88318924; PubMed=3412448;
RA Moos M., Tacke R., Scherer H., Teplov D., Frueh K., Schachner M.;
RT "Neural adhesion molecule L1 as a member of the immunoglobulin
RT superfamily with binding domain similar to fibronectin."
RL Nature 334:701-703(1988).
RL -1- FUNCTION: Cell adhesion molecule with an important role in the
development of the nervous system. Involved in neuron-neuron
adhesion, neurite fasciculation, outgrowth of neurites, etc. Binds
to axonin on neurons.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily.
CC L1/neurofascin/NG2CAM family.
CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 5 fibronectin type III domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.ebi.ac.uk/announcements
or send an email to license@ebi.ac.uk).
```

```

FT DOMAIN 1147 1260 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 35 130 IG-LIKE C2-TYPE 1.
FT DOMAIN 138 225 IG-LIKE C2-TYPE 2.
FT DOMAIN 239 337 IG-LIKE C2-TYPE 3.
FT DOMAIN 332 419 IG-LIKE C2-TYPE 4.
FT DOMAIN 424 506 IG-LIKE C2-TYPE 5.
FT DOMAIN 517 600 IG-LIKE C2-TYPE 6.
FT DOMAIN 827 896 FIBRONECTIN TYPE-III 1.
FT DOMAIN 932 994 FIBRONECTIN TYPE-III 2.
FT DOMAIN 1032 1094 FIBRONECTIN TYPE-III 3.
FT SITE 553 555 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 562 564 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 57 113 BY SIMILARITY.
FT DISULFID 157 208 BY SIMILARITY.
FT DISULFID 263 311 BY SIMILARITY.
FT DISULFID 353 403 BY SIMILARITY.
FT DISULFID 447 496 BY SIMILARITY.
FT DISULFID 538 590 BY SIMILARITY.
FT CARBOHYD 100 102 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 504 504 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 587 587 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 776 776 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 824 824 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 848 848 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 875 875 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 968 968 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 978 978 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1022 1022 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1030 1030 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1073 1073 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1107 1107 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1260 AA; 140968 MM; 22B857001CB2A558 CR64;

Query Match 4.2%; Score 144.5; DB 1; Length 1260;
Best Local Similarity 19.9%; Pred. No. 0.16;
Matches 118; Conservative 89; Mismatches 218; Indels 167; Gaps 30;

QY 34 GPTVELTCTASQKSIQFHWKNSNOIKLGNQSFITGPPSKLNDRAISRSLMDQGNFP 93
DB 346 GPTARLDQVQGRPOPEITWR-----INMSMETVKNQKRYRI-EGGS-- 387
QY 94 LIIKULKIEDSDTYICEVEDOK-----EEVQLVFGGLTANSDTHL-LOGQSLTLTL 143
DB 388 LILSNVQPTDVTWTOCEARNQGLLANNAYIVVQPARILTKYDQTVAAVGGSTAYLLC 447
QY 144 ESPSPSPSVQCRSPRGKNI-----QSGKTLVSQLELDQSGTWCTVLAQNKYEF 195
DB 448 KAFGARPVPSQWLDSEGTIVLQDERFFPYANNTLSTRDQANDTRGYFOANDQNTNTI 507
QY 136 KLDIVLAFQKASST-----VYKKEGEOVESFPPLAFTVEKLTSGSELIMWAQERASSSK 249
DB 508 LANLQV---KEATQITQGRSAIEKKGARVFTCOASFPSSL---QASITWRGDR----- 557
QY 250 SMITDILNKKEYSVRKVNQDPLQNGKKLPLHLTPQLPOVAGSGNLTLMLEATKGLKH 309
DB 558 -----DLQERGSDSKYFIJEDGLVTV-----QSL-DYSDQANSVCVASTLDEVE 600
QY 310 QEVNLVWMA-----TOLQKNLTCEWGP-----TSPKLMSLKENKEAK 350
DB 601 SRAQLLVGSSPGPVPHLELSDRLHLKQSQVHLS---WSPADBNHPLEKYEIDFEFDKE-- 655
QY 351 VSKREKPVAVLNPBEGMMOCLLSDGQV---LLESNIKVLPTWSTPVEBKSCDKTHTCPP 407
DB 656 -----MAPE---KMFSL-----GKVPNGNSTTLKLSPVHYVTFRTVAINKYGPSP 699

```

```

QY 408 CPAPELGCPSPVFLPPPKKDTLMISRPTEVCVVVDVSHDEPEVK-----FNW 456
DB 700 SPVSSVVTTPA-----APEK--PVDVRGSGENNNMNVITMKPLRMWDM 742
QY 457 YVDGEVHNNAKTPREQYNSYTRVSVLYTLHODMNGKCKVSKKALPAPIEKTIS 516
DB 743 NAPIQO-YRVQWRPQGGQ--ETWRKQTV-----SDPFLVSNSTSFVPEIKVQ 788
QY 517 KAKQGR--EPOV---YT-----LPRSDLEITKNQVSLTGLVNGFYSDIA 557
DB 789 AVNNQGKPEPQVYITIGYSGEDYPOVSPLELDTIFN-SSTVLVR-WRPVDLA 838

RESULT 98
KAC_MOUSE
ID_KAC_MOUSE STANDARD; PRT; 106 AA.
AC P01837;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig kappa chain C region.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE (MOPC 21).
RX MEDLINE=7305310; PubMed=4638343;
RA Svaeti J., Milstein C.;
RT "The complete amino acid sequence of a mouse kappa light chain.";
RL Biochem. J. 128:427-444(1972).
RN [2]
RP DISULFIDE BONDS (MOPC 21).
RX MEDLINE=73008889; PubMed=5073237;
RA Svaeti J., Milstein C.;
RT "The disulphide bridges of a mouse immunoglobulin G1 protein.";
RL Biochem. J. 126:837-850(1972).
RN [3]
RP SEQUENCE FROM N.A., AND REVISIONS TO 53-59.
RX MEDLINE=79084137; PubMed=103625;
RA Hamlyn P.H., Brownlee G.G., Cheng C.-C., Galt M.J., Milstein C.;
RT "Complete sequence of constant and 3' noncoding regions of an immunoglobulin mRNA using the dideoxynucleotide method of RNA sequencing.";
RL Cell 15:1067-1075(1978).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=82059477; PubMed=6170937;
RA Hamlyn P.H., Galt M.J., Milstein C.;
RT "Complete sequence of an immunoglobulin mRNA using specific priming and the dideoxynucleotide method of RNA sequencing.";
RL Nucleic Acids Res. 9:4485-4494(1981).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=81191915; PubMed=6262318;
RA Max E.E., Maizel J.V., Jr., Leder P.;
RT "The nucleotide sequence of a 5.5-kilobase DNA segment containing the mouse kappa immunoglobulin J and C region genes.";
RL J. Biol. Chem. 256:5116-5120(1981).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=81198949; PubMed=6785724;
RA Alendburger W., Neumaier P.S., Steinmetz M., Zachau H.G.;
RT "DNA sequence of the constant gene region of the mouse immunoglobulin kappa chain.";
RL Nucleic Acids Res. 9:971-981(1981).
RN [7]
RP SEQUENCE FROM N.A.
RX MEDLINE=88329081; PubMed=3138116;
RA de Waale P., Fey V., van de Voorde A., Mollemans F., Piers W.;
RT "Expression in non-Lymphoid cells of mouse recombinant immunoglobulin directed against the tumour marker human placental alkaline phosphatase.";

```

```

RL Eur. J. Biochem. 176:287-295 (1988).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; V00807; CA24189.1; -.
DR PIR; B90262; KIMS.
DR PDB; 1A1F; 01-FEB-97.
DR PDB; 1FSK; 02-OCT-00.
DR PDB; 1KB5; 08-APR-98.
DR PDB; 1KCR; 11-MAY-02.
DR PDB; 1KCS; 11-MAY-02.
DR PDB; 1KCU; 11-MAY-02.
DR PDB; 1KCV; 11-MAY-02.
DR PDB; 25C8; 09-JUL-99.
DR MGD; MGI:96495; IGK-C.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig1.
DR SMART; SM00407; Ig1.1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
DR Immunoglobulin domain; Immunoglobulin C region; 3D-structure.
FT NON_TER 1 1
FT DOMAIN 5 102 IG-LIKE.
FT DISULFID 26 86
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
SQ SEQUENCE 106 AA; 11778 MW; 4B51F5FE49B8E5 CEC64;

Query Match 4.2%; Score 143; DB 1; Length 106;
Best Local Similarity 31.5%; Pred. No. 0.0075;
Matches 29; Conservative 18; Mismatches 33; Indels 12; Gaps 2;

Oy 525 PQVYTLPSRDELTKNGVSLTCLVGFSPSDIAVEMENSGPENNYKTPPVLS----- 579
Db 5 PTVSIFPPSSQSLISGASVVCFLNNFYPKQINAVAKTKDSERON-----GVLSMTDQD 59

Oy 580 --DGSFFLYSKLTVDKSRKQGQVFNFSQVME 609
Db 60 SKDSTYSMSSTLTTLTKDEYERHNSYTCATHK 91

RESULT 99
NCM2_MOUSE STANDARD; PRT; 837 AA.
AC 035136; 035962;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neural cell adhesion molecule 2 precursor (N-CAM 2) (RB-8 neural cell
DE adhesion molecule) (R4B12).
GN NCAM2 OR OCAM OR RNCAM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RC STRAIN=BALB/c; TISSUE=Olfactory neuroepithelium;
RX MEDLINE=97368238; PubMed=9221781;
RA Yoshinaka Y., Kawasaki M., Tamada A., Fujita H., Hayashi H.,
RA Kagamiyama H., Mori K.;
RT "OCAM: A new member of the neural cell adhesion molecule family
RT related to zone-to-zone projection of olfactory and vomeronasal
RT axons.";
RL J. Neurosci. 17:5830-5842 (1997).

```

```

RN [2]
RP SEQUENCE FROM N.A. (ISOFORM SHORT).
RC STRAIN=C57BL/6J; TISSUE=Olfactory epithelium;
RX MEDLINE=97476194; PubMed=9334170;
RA Alenius M., Bohm S.;
RT "Identification of a novel neural cell adhesion molecule-related gene
RT with a potential role in selective axonal projection.";
RL J. Biol. Chem. 272:26083-26086 (1997).
CC -1- FUNCTION: May play important roles in selective fasciculation and
CC zone-to-zone projection of the primary olfactory axons.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (long isoform) and
CC attached to the membrane by a GPI-anchor (short isoform).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=O35136-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=O35136-2; Sequence=VSP_002590;
CC -1- TISSUE SPECIFICITY: Expressed in subsets of both olfactory and
CC vomeronasal neurons in a zone-specific manner.
CC -1- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 2 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF001287; AAB69125.1; -.
DR EMBL; AF001286; AAB69124.1; -.
DR EMBL; AF016619; AAC53375.1; -.
DR MGD; MGI:97282; Ncam2.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR Pfam; PF00047; Fn3_2.
DR Pfam; PF00047; Fn3_2.
DR SMART; SM00060; FN3_2.
DR SMART; SM00408; IGc2; 5.
DR PROSITE; PS50835; IG_LIKE; 5.
DR KEGG; K04901; Transmembrane; GPI-anchor; Alternative splicing.
KW Immunoglobulin domain; Signal; GPI-anchor; Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 837
FT DOMAIN 20 697
FT TRANSMEM 698 718
FT DOMAIN 719 837
FT DOMAIN 21 108
FT DOMAIN 113 202
FT DOMAIN 208 297
FT DOMAIN 302 396
FT DOMAIN 401 491
FT DOMAIN 482 581
FT DOMAIN 594 678
FT DISULFID 42 93
FT DISULFID 136 186
FT DISULFID 232 281
FT DISULFID 322 380
FT DISULFID 422 475
FT CARBOHYD 177 177
FT CARBOHYD 219 219
FT CARBOHYD 309 309
FT CARBOHYD 406 406
FT CARBOHYD 419 419
FT CARBOHYD 445 445
FT CARBOHYD 474 474
FT CARBOHYD 562 562
FT VARSPLIC 694 837

```



```

FT      DEBITNHDGSPVNEBETPTPLTEPEKCLPLKEENGKEVLNA
FT      ETEIKVNDIIOCKEDDIKA -> NCCGANGGNGGSMH
FT      LNAVGTPTVITMSISLCLF (in isoform short).
FT      /FTid=VSP_002590.
SQ      SEQUENCE      837 AA; 93203 MW; 70473B053A2D65A5 CRC64;
Query Match      4.2%; Score 142; DB 1; Length 837;
Best Local Similarity 17.7%; Pred. No. 0.13;
Matches 101; Conservative 87; Mismatches 210; Indels 174; Gaps 22;

QY      3 RG-VPEPHLLVLTQALLPRAATQGNKRVLAGKGDVLTCTASQKKSIOGFHKNSNQIKI 61
DB      193 REIDPRDIIIVYVNPRAIMPPQKSNFATARGEEMTLTCKASGSPDPTISFRNG--KL 250
QY      62 LQNGSFLTKGPSKLNDRADSRSLMDQGNPFIITNKLIEDSDTYICEV-----EDQKE 116
DB      251 ISENEKYIILKG-----SNTLETVNINIKKGGSVVCATNKAGEDQKQ 293
QY      117 E-VOLLVFGLTANSPTHLQOGS-----LTLTLESPGPSSPVQC----- 155
DB      294 AFLQVAV-----QPHIIQLKNETTSENGHTLVCEAGEFVPEITMKRAIDGVWFSG 346
QY      156 -RSPRG---KNIOGGKTLVSQLELDGSGTWCTVLQ---NOKKVEFKIDIYVLAPOK 206
DB      347 DKSPGRIEVKGQGHSSLIHRDVKLSDSGRYDCEASRIIGGHQREMHLDIEYAP-KFVS 405
QY      207 ASSIYVKEGEOVERSPPLAFVVEKLTSGGELMWQAEKASSSSKSWITFDLKNKEVSVKV 266
DB      406 NOTWYMSWEGNPINISCDVT-----ANPPASIHMRREKLL-----LPAKNT 446
QY      267 TDDPKLQMGKLLPLHLTLPOALPOVAGSGNLTALAEAKTGKLNHOEVNLVVMBATQLOKML 326
DB      447 TLTKHTSVKRMKILEA-PTSDNDF---GRYNCTATNRIGTFOEYITELADV----- 495
QY      327 TCEVNGPTSPKMLSLKLENKEAKVSKREKPVVNLPEAGMOCCLSDSGVLLBSNIVK 386
DB      496 -----PSSPHGVKIIIEISQTAKISFNK----- 518
QY      387 LPTWSPVPEKSGCDKHTHTCPCPAPRLDGPVSFLFPPPKY---DTLMISTREYTCV 442
DB      519 -----PESHGCVPIHHYQVDVKEVASETWKIVSHGQTVV 554
QY      443 VDVSHDEPEVKFNWYDVGVVHNAKTKPREEOGNSYRVVSVLTV-----LHQDWMLG 495
DB      555 V-LSSLEPPTY-----EIRVAANGGQGDYSKIEIFQTLVPAEPPSPHIGQPSGG 606
QY      496 KEYKCVSNK-ALPAPIEKTISKAKGQPREPQ 526
DB      607 KSPKISITKQDDGCAPILEYIVKYRSKDEDDQ 638

RESULT 100
NCAM2_HUMAN      STANDARD; PRT; 761 AA.
AC      P13592; P13593;
DT      01-JAN-1990 (Rel. 13, Created)
DT      01-APR-1990 (Rel. 14, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Neural cell adhesion molecule 1, 120 kDa isoform precursor (N-CAM 120)
DE      (NCAM-120) (CD56 antigen).
GN      NCAM1 OR NCAM.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A. (ISOFORM N-CAM 120).
RC      TISSUE=Skeletal muscle;
RX      MEDLINE=89305258; PubMed=3253057;
RA      Barton C.H., Dickson G., Gower H.J., Rowlett L.H., Putt W.,
RA      Elsom V., Moore S.E., Goriadis C., Walsh F.S.;
RA      "Complete sequence and in vitro expression of a tissue-specific
RT      phosphatidylinositol-linked N-CAM isoform from skeletal muscle.",

```

```

RL      Development 104:165-173(1988).
RN      [2]
RP      SEQUENCE OF 491-761 FROM N.A. (ISOFORM N-CAM 120).
RC      TISSUE=Skeletal muscle;
RX      MEDLINE=87301755; PubMed=2887295;
RA      Dickson G., Gower H.J., Barton C.H., Prentice H.M., Elsom V.L.,
RA      Moore S.E., Cox R.D., Quinn C., Putt W., Walsh F.S.;
RA      "Human muscle neural cell adhesion molecule (N-CAM): identification
RT      of a muscle-specific sequence in the extracellular domain.";
RL      Cell 50:1119-1130(1987).
RN      [3]
RP      SEQUENCE OF 491-655 FROM N.A. (ISOFORM C).
RX      MEDLINE=89077552; PubMed=3203385;
RA      Gower H.J., Barton C.H., Elsom V.L., Thompson J., Moore S.E.,
RA      Dickson G., Walsh F.S.;
RT      "Alternative splicing generates a secreted form of N-CAM in muscle
RT      and brain.";
RL      Cell 55:955-964(1988).
CC      -!- FUNCTION: This protein is a cell adhesion molecule involved in
CC      neuron-neuron adhesion, neurite fasciculation, outgrowth of
CC      neurites, etc.
CC      -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC      -!- ALTERNATIVE PRODUCTS:
CC      Event=Alternative splicing; Named isoforms=3;
CC      Name=N-CAM 120;
CC      IsoId=P13592-2; Sequence=Displayed;
CC      Name=N-CAM 140;
CC      IsoId=P13591-1; Sequence=External;
CC      Name=C; Synonyms=Secreted;
CC      IsoId=P13592-1; Sequence=VSP_002587;
CC      -!- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.
CC      -!- SIMILARITY: Contains 2 fibronectin type III domains.
CC      -!- DATABASE: NAME=PROV; NOTE=CD guide CD56 entry;
CC      WWW="http://www.ncbi.nlm.nih.gov/ncbi/cd/cd56.htm".
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb.ch/announce/
CC      or send an email to license@isb.ch).
CC      -----
DR      EMBL: X16841; CAA34739.1; -
DR      EMBL: M17409; AAA59912.1; -
DR      EMBL: M22094; AAA59910.1; -
DR      EMBL: M22092; AAA59911.1; -
DR      EMBL: M22091; AAA59911.1; JOINED.
DR      PIR: A31635; A31635.
DR      PIR: S07784; IGHUNG.
DR      Genew; HGNC:7656; NCAM1.
DR      MIM; 116930; -.
DR      GO; GO:0016021; C:integral to membrane; TAS.
DR      GO; GO:0005865; C:plasma membrane; TAS.
DR      InterPro; IPR008957; FN_III-like.
DR      InterPro; IPR003961; FN_III.
DR      InterPro; IPR007110; IG-like.
DR      InterPro; IPR003598; IG_c2.
DR      Pfam; PF00041; fn3; 2.
DR      Pfam; PF00047; Ig; 5.
DR      SMART; SM00060; FN3; 2.
DR      SMART; SM00408; IGc2; 5.
DR      PROSITE; PS50835; IG-LIKE; 5.
KW      Immunoglobulin domain; Cell adhesion; Glycoprotein; Repeat; signal;
KW      GPI-anchor; Alternative splicing.
FT      SIGNAL      1 19
FT      CHAIN      20 761      NEURAL CELL ADHESION MOLECULE 1, 120 kDa
FT      ISOFORM.
FT      DOMAIN      20 111      IG-LIKE C2-TYPE 1.
FT      DOMAIN      116 205      IG-LIKE C2-TYPE 2.
FT      DOMAIN      212 301      IG-LIKE C2-TYPE 3.
FT      DOMAIN      308 403      IG-LIKE C2-TYPE 4.
FT      DOMAIN      406 491      IG-LIKE C2-TYPE 5.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 12:51:44 ; Search time 44.4955 Seconds
(without alignments)
4594.975 Million cell updates/sec

Title: SEQ3
Perfect score: 3414
Sequence: 1 NMRGVFRRHLVLQLALLP.....DETCAADQGEJLDTTDP 648

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 125 summaries

Database :
1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORNITHINE:*
9: SP_PLANT:*
10: SP_VIRUS:*
11: SP_VIRUS:*
12: SP_VIRUS:*
13: SP_VIRUS:*
14: SP_VIRUS:*
15: SP_VIRUS:*
16: SP_VIRUS:*
17: SP_VIRUS:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1714	50.2	397	6	009261 cercopithec
2	1710	50.1	397	6	009260 cercopithec
3	1709	50.1	397	6	009259 cercopithec
4	1705	49.9	397	6	009259 cercopithec
5	1705	49.9	397	6	009263 cercopithec
6	1698	49.7	397	6	009262 cercopithec
7	1697	49.7	397	6	009262 cercopithec
8	1610.5	46.4	457	6	008278 cercopithec
9	1584.5	46.4	457	6	008278 cercopithec
10	1277	37.4	470	4	0725W1 homo sapien
11	1266	37.1	469	4	0727P5 homo sapien
12	1265	37.1	509	4	08NP17 homo sapien
13	1264.5	37.0	482	4	072351 homo sapien
14	1258	36.8	479	4	096P08 homo sapien
15	1185.5	34.7	455	6	09XS78 delphinapte
16	1171	34.3	354	4	086TT2 homo sapien

17	1167	34.2	521	4	08N4Y9
18	1150	33.7	473	4	08TC63
19	1087	31.8	474	6	P79355
20	985	28.9	457	11	061396
21	957	28.0	433	11	055054
22	919	26.9	337	6	095M34
23	857	25.1	470	11	07TMR1
24	850.5	24.9	437	11	09R1A4
25	850.5	24.9	469	11	08R1V9
26	844	24.7	463	11	099LC4
27	840.5	24.6	473	11	099DL4
28	832	24.4	468	11	099LJ1
29	810.5	23.7	473	11	099LZ5
30	800.5	23.4	473	11	091Z05
31	799.5	23.4	474	11	08R3H6
32	417.5	12.2	613	4	096EY0
33	416	12.2	614	11	07TMT6
34	407.5	11.9	613	4	08WUK1
35	400	11.7	86	6	077596
36	400	11.7	86	6	077597
37	397	11.6	86	6	077594
38	397	11.6	86	6	077599
39	396.5	11.6	613	11	08VCX7
40	395	11.6	86	6	077595
41	392.5	11.5	597	4	096BB9
42	392	11.5	86	6	077598
43	389	11.4	614	4	096GAC
44	388	11.4	86	6	077601
45	383	11.2	618	6	077600
46	383	11.2	618	4	096AA6
47	378.5	11.1	597	4	09BUI0
48	376.5	11.0	597	4	09BUI8
49	373.5	10.9	588	4	08WUX4
50	357.5	10.5	375	4	086T11
51	357	10.5	71	4	013969
52	333.5	9.8	587	13	07TOR1
53	330.5	9.7	684	13	090544
54	305.5	8.9	99	6	029027
55	302.5	8.9	478	4	072379
56	298.5	8.7	494	4	096K68
57	296	8.7	496	4	096KX8
58	295	8.6	499	4	08NSK4
59	292	8.6	493	4	08NCL6
60	291	8.5	416	4	09NPP6
61	288	8.4	500	4	09BRV0
62	287.5	8.4	496	4	096DK0
63	286	8.4	492	4	072374
64	282	8.3	488	11	091WR1
65	281.5	8.2	384	4	09UP60
66	280.5	8.2	99	6	029028
67	279.5	8.2	497	4	08WY24
68	276.5	8.1	487	11	080217
69	274.5	8.0	487	11	099KX4
70	273	8.0	120	11	088650
71	272.5	8.0	489	11	08VCX4
72	271.5	8.0	482	11	090WB5
73	271.5	8.0	484	11	099LA6
74	270.5	7.9	486	11	091Z07
75	270	7.9	481	11	091W17
76	270	7.9	482	11	08K172
77	269.5	7.9	426	11	09DCD9
78	266	7.8	481	11	091WT3
79	266	7.8	482	11	091W32
80	265.5	7.8	479	11	091W55
81	264	7.7	484	11	08VEA0
82	264	7.7	488	11	08KOP2
83	261	7.6	480	11	091XEL
84	255.5	7.5	481	11	08VCV5
85	252	7.4	479	11	07TMR4
86	251.5	7.4	480	11	08K0Z4
87	250.5	7.3	479	11	099M22
88	240	7.0	487	13	09W6V7
89	237	6.9	90	11	P70443

08N4Y9	homo sapien
08TC63	homo sapien
P79355	felis silve
061396	mus musculus
055054	mus musculus
095M34	equus caball
07TMR1	mus musculus
09R1A4	mus musculus
08R1V9	mus musculus
099LC4	mus musculus
099DL4	mus musculus
099LJ1	mus musculus
099LZ5	mus musculus
091Z05	mus musculus
08R3H6	mus musculus
096EY0	homo sapien
07TMT6	mus musculus
08WUK1	homo sapien
077596	mandrillus
077597	mandrillus
077594	cercopithec
077599	theropithec
08VCX7	mus musculus
077595	cercopithec
096BB9	homo sapien
077598	papio sp. (
096GAC	homo sapien
077601	lophocebus
077600	lophocebus
096AA6	homo sapien
09BUI0	homo sapien
09BUI8	homo sapien
08WUX4	homo sapien
086T11	homo sapien
013969	homo sapien
07TOR1	xenopus lae
090544	ginglymocto
029027	sus scrofa
072379	homo sapien
096K68	homo sapien
096KX8	homo sapien
08NSK4	homo sapien
08NCL6	homo sapien
09NPP6	homo sapien
09BRV0	homo sapien
096DK0	homo sapien
072374	homo sapien
091WR1	mus musculus
09UP60	homo sapien
029028	sus scrofa
08WY24	homo sapien
080217	mus musculus
099KX4	mus musculus
088650	marmota mon
08VCX4	mus musculus
090WB5	anas platyr
099LA6	mus musculus
091Z07	mus musculus
091W17	mus musculus
08K172	mus musculus
09DCD9	mus musculus
091WT3	mus musculus
091W32	mus musculus
091W55	mus musculus
08VEA0	mus musculus
08KOP2	mus musculus
091XEL	mus musculus
08VCV5	mus musculus
07TMR4	mus musculus
08K0Z4	mus musculus
099M22	mus musculus
09W6V7	gallus gall
P70443	mus musculus

```

90 190 5.6 5636 4 Q96RW7
91 186 5.4 268 13 Q90524
92 184.5 5.4 573 4 Q8WU38
93 183.5 5.4 130 11 Q9D8W4
94 183 5.4 234 4 Q8N355
95 182.5 5.3 739 6 Q28260
96 182.5 5.3 739 11 Q63669
97 182 5.3 17352 5 Q95YMW2
98 180.5 5.3 2693 5 Q8ISF3
99 180.5 5.3 2708 5 Q8ISF4
100 180.5 5.3 2808 5 Q8MNS0
101 180.5 5.3 18519 5 Q8ISF6
102 180.5 5.3 18519 5 Q8ISF6
103 180.5 5.3 18534 5 Q8ISF7
104 180 5.3 235 11 Q99M11
105 179 5.2 1031 13 Q90YMW2
106 178.5 5.2 2673 4 Q96SC3
107 178.5 5.2 6620 4 Q96AA2
108 177.5 5.2 4162 13 Q98918
109 176.5 5.2 233 4 Q96169
110 176.5 5.2 259 13 Q90530
111 176.5 5.2 1323 13 Q08476
112 175.5 5.1 2053 4 Q8WU37
113 175.5 5.1 2113 4 Q8TD84
114 174.5 5.1 8081 5 Q72120
115 174 5.1 236 4 Q96E61
116 174 5.1 650 6 Q96KR2
117 174 5.1 739 6 Q96KR2
118 173 5.1 34350 4 Q8W242
119 172.5 5.1 234 4 Q722U7
120 172.5 5.1 2053 4 Q81ZV4
121 172.5 5.1 7962 4 Q10465
122 171.5 5.0 240 4 Q8WU33
123 171.5 5.0 257 13 Q90536
124 171 5.0 233 4 Q8N5F4
125 170 5.0 237 13 Q90545

```

ALIGNMENTS

```

RESULT 1
ID 009261 PRELIMINARY; PRT; 397 AA.
AC 009261;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE CD4 (Fragment).
GN CD4.
OS Cercopithecus sabaenus.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Cercopithecus.
CX NCBI_TaxID=60711;
RN NCBI_TaxID=60711;
RP SEQUENCE FROM N.A.
RX MEDLINE=98017879; PubMed=9379478;
RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA Corbet S., Barre-Sinoussi F., Allan J.S.;
RA "Relation between phylogeny of African green monkey CD4 genes and
RA their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
DR EMBL; AF001225; AAB60872.1; -
DR HSSP; P01730; IWIQ.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR000973; CD4_TCSG.
DR InterPro; IPR007110; IG_Like.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR SMART; SM00406; IGV; 1.

```

```

DR PROSITE; PSS0835; IG_Like; 1.
FT NON_TER 1
FT NON_TER 397
SQ SEQUENCE 397 AA; 43945 MW; 80C963B92A68CD3 CRC64;
Query Match 50.2%; Score 1714; DB 6; Length 397;
Best Local Similarly 89.7%; Pred. No. 3,6e-122;
Matches 331; Conservative 16; Mismatches 22; Indels 0; Gaps 0;
QY 28 VLKGGKDTVELTCTGSOKSIQPHMKNNOIKILGQSGFLTKGSPKLNDRADRSRLW 87
D 1 VLKGGKDTVELTCTGNSQNTTTPHMKNSQIKILGQSGFLTKGSKLNDRIDRSKSLW 60
QY 88 DQGNFPLIKNLKIEDSDTYICEVEDQKEVOLVFLGLTANSPTHLAQGSLTTLTSSPP 147
D 61 DQGFSSMIINLNLKIEDSEYICEVENKEVELLVFLTANSPTHLAQGSLTTLTSSPP 120
QY 148 GSSPSVOCRRSPRGKNTGGKTLTSSQLELSDSGTWTTLVQNKQKVEFKIDIVVLAQKA 207
D 121 GSSPSVCRSPRGKNTGGKTLTSSVQLERODSGTWTCTVSDQNMVEFKIDIVVLAQKA 180
QY 208 SSIYKKKGROVERSPFLATFVEKLTGSGELMWQAERASSKSWITFDLNKKEVSVRVT 267
D 181 SSIYKKKGROVERSPFLATFVEKLTGSGELMWQAERASSKSWITFDLNKKEVSVRVT 240
QY 268 QDPRLQWKGKPLPLTLTPOALPOYAGSGLTLALAEAKTGLHDEVNLVWRATQQLNT 327
D 241 QDPRLQWKGKPLPLTLTPOALPOYAGSGLTLALAEAKTGLHDEVNLVWRATQQLNT 300
QY 328 CEWGPSPKLMISLKENKAESKREKPVWVLANPAGMWQCLISGGOVLLESNIKVL 387
D 301 CEWGPSPKLMISLKENKAATVSKAKAVVLANPAGMWQCLISGGOVLLESNIKVL 360
QY 388 PTWSTPVEP 396
D 361 PTWSTPVEP 369

```

```

RESULT 2
ID 009260 PRELIMINARY; PRT; 397 AA.
AC 009260;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE CD4 (Fragment).
GN CD4.
OS Cercopithecus sabaenus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Cercopithecus.
CX NCBI_TaxID=60711;
RN NCBI_TaxID=60711;
RP SEQUENCE FROM N.A.
RX MEDLINE=98017879; PubMed=9379478;
RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA Corbet S., Barre-Sinoussi F., Allan J.S.;
RA "Relation between phylogeny of African green monkey CD4 genes and
RA their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
DR EMBL; AF001224; AAB60871.1; -
DR HSSP; P01730; IWIQ.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR000973; CD4_TCSG.
DR InterPro; IPR007110; IG_Like.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PSS0835; IG_Like; 1.
FT NON_TER 1
FT NON_TER 397

```

SQ SEQUENCE 397 AA; 43882 MW; 478BB27E992EE89 CRC64;
 Query Match 50.1%; Score 1710; DB 6; Length 397;
 Best Local Similarity 89.4%; Pred. No. 7.3e-122;
 Matches 330; Conservative 17; Mismatches 22; Indels 0; Gaps 0;

QY 28 VVLGKKGDVLTCTASQKSIQPHWKNNOIKILGNQGSFLTKGPKLNDRAISRSLW 87
 1 VVLGKKGDVLTCTASQKSIQPHWKNNOIKILGNQGSFLTKGPKLNDRAISRSLW 60
 DB
 QY 88 DQGNPFLIKNLKIEDSDTYICEVEDQKEVQLVFGLTANSDDLLOGOSITLTLESPP 147
 61 DQGCSSMIINLKNKIEDSEITYICEVENKEVELVFGLTANSDDLLOGOSITLTLESPP 120
 DB
 QY 148 GSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLONQKVEFKIDIVLAFOKA 207
 121 GSSPSVKCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLONQKVEFKIDIVLAFOKA 180
 DB
 QY 208 SSIYVKKGEQVEFSPFLAFVTEKLTGSGELMWQERASSSSKSWITFDLKNKEVSKVKT 267
 181 SSIYVKKGEQVEFSPFLAFVTEKLTGSGELMWQERASSSSKSWITFDLKNKEVSKVKT 240
 DB
 QY 268 QDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALAEKTKGLHQBVLVVMRATOLQKILT 327
 241 QDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALAEKTKGLHQBVLVVMRATOLQKILT 300
 DB
 QY 328 CEVWGPTSPKMLSLKLENKEAKVSKREKPVWVLPNPEAGMOCILSDSGVLLSNIKYL 387
 301 CEVWGPTSPKMLSLKLENKEAKVSKREKPVWVLPNPEAGMOCILSDSGVLLSNIKYL 360
 DB
 QY 388 PTWSTPVEP 396
 361 PTWSTPVEP 369
 DB

RESULT 3
 095NE9 PRELIMINARY; PRT; 397 AA.
 AC 095NE9;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE CD4 (Fragment).
 GN CD4.
 OS Cercopithecus pygerythrus.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
 OC Cercopithecinae; Cercopithecus.
 OX NCBI_TaxId=60710;
 RX MEDLINE=98017679; PubMed=9379478;
 RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
 Corbet S., Barre-Sinoussi F., Allan J.S.;
 RT "Relation between phylogeny of African green monkey CD4 genes and
 their respective simian immunodeficiency virus genes.";
 RL J. Med. Primatol. 26:120-128(1997).
 DR EMBL; AF001223; AAB60870.1; -.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0006955; P:immune response; IEA.
 DR InterPro; IPR000973; CD4_TcAg.
 DR InterPro; IPR007110; Ig_Like.
 DR InterPro; IPR003596; Ig_v.
 DR Pfam; PF00047; Ig_2.
 DR PRINTS; PR00692; CD4TCANTIGEN.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PSS0835; IG_LIKE; 1.
 FT NON_TER 1
 FT NON_TER 1
 SQ SEQUENCE 397 AA; 43946 MW; 21CE30882ABFBC0 CRC64;

Query Match 50.1%; Score 1709; DB 6; Length 397;
 Best Local Similarity 89.4%; Pred. No. 8.7e-122;

Matches 330; Conservative 16; Mismatches 23; Indels 0; Gaps 0;

QY 28 VVLGKKGDVLTCTASQKSIQPHWKNNOIKILGNQGSFLTKGPKLNDRAISRSLW 87
 1 VVLGKKGDVLTCTASQKSIQPHWKNNOIKILGNQGSFLTKGPKLNDRAISRSLW 60
 DB
 QY 88 DQGNPFLIKNLKIEDSDTYICEVEDQKEVQLVFGLTANSDDLLOGOSITLTLESPP 147
 61 DQGCSSMIINLKNKIEDSEITYICEVENKEVELVFGLTANSDDLLOGOSITLTLESPP 120
 DB
 QY 148 GSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLONQKVEFKIDIVLAFOKA 207
 121 GSSPSVKCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLONQKVEFKIDIVLAFOKA 180
 DB
 QY 208 SSIYVKKGEQVEFSPFLAFVTEKLTGSGELMWQERASSSSKSWITFDLKNKEVSKVKT 267
 181 SSIYVKKGEQVEFSPFLAFVTEKLTGSGELMWQERASSSSKSWITFDLKNKEVSKVKT 240
 DB
 QY 268 QDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALAEKTKGLHQBVLVVMRATOLQKILT 327
 241 QDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALAEKTKGLHQBVLVVMRATOLQKILT 300
 DB
 QY 328 CEVWGPTSPKMLSLKLENKEAKVSKREKPVWVLPNPEAGMOCILSDSGVLLSNIKYL 387
 301 CEVWGPTSPKMLSLKLENKEAKVSKREKPVWVLPNPEAGMOCILSDSGVLLSNIKYL 360
 DB
 QY 388 PTWSTPVEP 396
 361 PTWSTPVEP 369
 DB

RESULT 4
 009259 PRELIMINARY; PRT; 397 AA.
 AC 009259;
 DT 01-JUL-1997 (TREMBLrel. 04, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE CD4 (Fragment).
 GN CD4.
 OS Cercopithecus sabaues.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
 OC Cercopithecinae; Cercopithecus.
 OX NCBI_TaxId=60711;
 RX MEDLINE=98017879; PubMed=9379478;
 RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
 Corbet S., Barre-Sinoussi F., Allan J.S.;
 RT "Relation between phylogeny of African green monkey CD4 genes and
 their respective simian immunodeficiency virus genes.";
 RL J. Med. Primatol. 26:120-128(1997).
 DR EMBL; AF001223; AAB60870.1; -.
 DR HSP; P01730; 1WIO.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0006955; P:immune response; IEA.
 DR InterPro; IPR00973; CD4_TcAg.
 DR InterPro; IPR007110; Ig_Like.
 DR InterPro; IPR003596; Ig_v.
 DR Pfam; PF00047; Ig_2.
 DR PRINTS; PR00692; CD4TCANTIGEN.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PSS0835; IG_LIKE; 1.
 FT NON_TER 1
 FT NON_TER 1
 SQ SEQUENCE 397 AA; 43881 MW; 7CE39AD0F8506C81 CRC64;

Query Match 49.9%; Score 1705; DB 6; Length 397;
 Best Local Similarity 89.2%; Pred. No. 1.8e-121;
 Matches 329; Conservative 18; Mismatches 22; Indels 0; Gaps 0;

QY 28 VVLGKKGDVLTCTASQKSIQPHWKNNOIKILGNQGSFLTKGPKLNDRAISRSLW 87

```
Db 1 VVLGKGDVLTCTASQKTSIOFHWNKSNQIKILGQGSFLTKGSKLRIDRSKSLW 60
Qy 88 DQGNFPLIKNLKIEDSDTYICEVEDOKEEVOALLVFGLTANSPTHLQGGSLTLTLESPP 147
Db 61 DQGFMSWIIKNLKIEDSETYICEVENKEEVEALLVFGLTANSPTHLQGGSLTLTLESPP 120
Qy 148 GSSPSVQCRSPRGKNIQGGKTLISVSQLBLQDSGTWCTTVLQNKKEVFKIDIVVLAFOKA 207
Db 121 GSSPSVQCRSPRGKNIQGGKTLISVSQLBLQDSGTWCTTVLQNKKEVFKIDIVVLAFOKA 180
Qy 208 SSIYKKEGQVEFSFPLATFVEKLTGSGELMWOAERASSSKSWITFDLKNKEVSVKRYT 267
Db 181 SSIYKKEGQVEFSFPLATFVEKLTGSGELMWOAERASSSKSWITFDLKNKEVSVKRYT 240
Qy 268 QDPKLQWKKLPLNLTLPQALPYAGSGNLTLLAEAKTGLHDEVNLVWMRATQFQENLT 327
Db 241 QDPKLQWKKLPLNLTLPQALPYAGSGNLTLLAEAKTGLHDEVNLVWMRATQFQENLT 300
Qy 328 CEVWGPTSPKLMSTKLENKEAKYSKREKPVWLNPEAGMOCCLSDSGVLLSNIKVL 387
Db 301 CEVWGPTSPKLMSTKLENKEAKYSKREKPVWLNPEAGMOCCLSDSGVLLSNIKVL 360
Qy 388 PTWSTPVEP 396
Db 361 PTWPTPVQP 369
```

```
RESULT 5
ID 009263 PRELIMINARY; PRT; 397 AA.
AC 009263;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE CD4 (Fragment).
GN CD4.
OS Cercopithecus tantalus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=60712;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98017879; PubMed=9379478;
RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA Corbet S., Barre-Sinoussi F., Allan J.S.;
RT "Relation between phylogeny of African green monkey CD4 genes and
RT their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
DR EMBL; AF001222; AAB60869.1; -.
DR HSSP; P01730; IWIQ.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig_Like.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 397
SQ SEQUENCE 397 AA; 43994 MW; A3CD031535A51524 CRC64;
```

Query Match 49.9%; Score 1705; DB 6; Length 397;
Best Local Similarity 89.2%; Pred. No. 1.8e-121;
Matches 329; Conservative 16; Mismatches 24; Indels 0; Gaps 0;

Qy 28 VVLGKGDVLTCTASQKTSIOFHWNKSNQIKILGQGSFLTKGSKLRIDRSKSLW 87
Db 1 VVLGKGDVLTCTASQKTSIOFHWNKSNQIKILGQGSFLTKGSKLRIDRSKSLW 60

```
Qy 88 DQGNFPLIKNLKIEDSDTYICEVEDOKEEVOALLVFGLTANSPTHLQGGSLTLTLESPP 147
Db 61 DQGFMSWIIKNLKIEDSETYICEVENKEEVEALLVFGLTANSPTHLQGGSLTLTLESPP 120
Qy 148 GSSPSVQCRSPRGKNIQGGKTLISVSQLBLQDSGTWCTTVLQNKKEVFKIDIVVLAFOKA 207
Db 121 GSSPSVQCRSPRGKNIQGGKTLISVSQLBLQDSGTWCTTVLQNKKEVFKIDIVVLAFOKA 180
Qy 208 SSIYKKEGQVEFSFPLATFVEKLTGSGELMWOAERASSSKSWITFDLKNKEVSVKRYT 267
Db 181 SSIYKKEGQVEFSFPLATFVEKLTGSGELMWOAERASSSKSWITFDLKNKEVSVKRYT 240
Qy 268 QDPKLQWKKLPLNLTLPQALPYAGSGNLTLLAEAKTGLHDEVNLVWMRATQFQENLT 327
Db 241 QDPKLQWKKLPLNLTLPQALPYAGSGNLTLLAEAKTGLHDEVNLVWMRATQFQENLT 300
Qy 328 CEVWGPTSPKLMSTKLENKEAKYSKREKPVWLNPEAGMOCCLSDSGVLLSNIKVL 387
Db 301 CEVWGPTSPKLMSTKLENKEAKYSKREKPVWLNPEAGMOCCLSDSGVLLSNIKVL 360
Qy 388 PTWSTPVEP 396
Db 361 PTWPTPVQP 369
```

```
RESULT 6
ID 009262 PRELIMINARY; PRT; 397 AA.
AC 009262;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE CD4 (Fragment).
GN CD4.
OS Cercopithecus tantalus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=60712;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98017879; PubMed=9379478;
RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA Corbet S., Barre-Sinoussi F., Allan J.S.;
RT "Relation between phylogeny of African green monkey CD4 genes and
RT their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
DR EMBL; AF001221; AAB60868.1; -.
DR HSSP; P01730; IWIQ.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 397
SQ SEQUENCE 397 AA; 43954 MW; CF7F2F5D82335B0D CRC64;
```

Query Match 49.7%; Score 1698; DB 6; Length 397;
Best Local Similarity 89.2%; Pred. No. 6e-121;
Matches 329; Conservative 16; Mismatches 24; Indels 0; Gaps 0;

Qy 28 VVLGKGDVLTCTASQKTSIOFHWNKSNQIKILGQGSFLTKGSKLRIDRSKSLW 87
Db 1 VVLGKGDVLTCTASQKTSIOFHWNKSNQIKILGQGSFLTKGSKLRIDRSKSLW 60
Qy 88 DQGNFPLIKNLKIEDSDTYICEVEDOKEEVOALLVFGLTANSPTHLQGGSLTLTLESPP 147
Db 61 DQGFMSWIIKNLKIEDSETYICEVENKEEVEALLVFGLTANSPTHLQGGSLTLTLESPP 120

```

QY 148 GSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTCTCTVLQONOKVYEFKIDIVLAFQKA 207
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 GSSPSVKCSPRGKNIQGGRTLSVPLEERQDSGTCTCTVSQDNVYEFKIDIVLAFQKA 180
QY 208 SSIYVKEGEQVEFSFPLAFVTEKLTGSGELVMQWQERASSSKSWITFDLKNKEVSKVKT 267
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 SSTYVKEGEQVEFSFPLAFVTEKLTGSGELRMQWQERASSSKSWITFDLKNKEVSKVKT 240
QY 268 QDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALAEKTKLHQEVNLVVMRATQLOKNT 327
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 QDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALAEKTKLHQEVNLVVMRATQLOKNT 300
QY 328 CEVWGPTSPKMLSLKLENKKAIVSKREKPVWVNLNPEAGMOCCLSDSGQVLLSNIKVL 387
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 CEVWGPTSPKMLSLKLENKKAIVSKQAKAVWVNLNPEAGMOCCLSDSGQVLLSNIKVL 360
QY 388 PTWSTPVP 396
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 PTWSTPVP 369

```

RESULT 7

```

ID 002805 PRELIMINARY; PRT; 397 AA.
AC 002805; 07593;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
    (Fragment).
GN CD4.
OS Cercopithecus aethiops (Green monkey) (Grivet).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OC NCBI_TaxID=9534;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PerIPHERAL blood;
RA MEDLINE=98017879; PubMed=9379478;
RA Fomsgaard A., Mueller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA Corbet S., Barre-Sinoussi F., Allan J.S.;
RA "Rejection between phylogeny of African green monkey CD4 genes and
RT their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
RN [2]
RP SEQUENCE OF 80-165 FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.E., Diocelli T.R.;
RA "Nuclear gene trees and the phylogenetic relationships of the
RT mangabeys (primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH p56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 3 C2-LIKE DOMAINS.
DR EMBL; AF001226; AAB60873.1; -
DR EMBL; AF057380; AAC25124.1; -
DR HSBP; P01730; IWI0.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR000973; CD4 TCAG.
DR InterPro; IPR007110; IG-Like.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF000447; I9; 2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Lipoprotein; Palmitate; Repeat.

```

```

FT NON TER 1 1
FT DOMAIN <1 370
FT TRASMEM 371 391
FT DOMAIN 392 >397
FT DOMAIN <1 98
FT DOMAIN 99 176
FT DOMAIN 177 290
FT DOMAIN 291 347
FT CAROHD 15 15
FT CAROHD 30 30
FT CAROHD 269 269
FT CAROHD 298 298
FT DISULFD 14 82
FT DISULFD 128 157
FT DISULFD 301 343
FT LIPID 392 392
FT LIPID 395 395
FT NON TER 397 397
SQ SEQUENCE 397 AA; 43980 MW; F74C42E22B196155 CRC64;

```

Query Match 49.7%; Score 1697; DB 6; Length 397;
 Best Local Similarity 88.9%; Pred. No. 7.2e-121;
 Matches 328; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

```

QY 28 VVLGKKGPVVELTCTASQKSIQFMKNSNOIKILGNQSFLLTKGSPKLNDRADSRSLM 87
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 VVLGKKGPVVELTCTASQKTIQFMKNSNOIKILGKQSFLLTKGSKLRIDSRKSLM 60
QY 88 DQGNPLIINKLKIEDSDTYICEVEDQEEVQLVFGLTANSDDLQGSYTLTLSP 147
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 DQGNPLIINKLKIEDSDTYICEVEDQEEVQLVFGLTANSDDLQGSYTLTLSP 120
QY 148 GSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTCTCTVLQONOKVYEFKIDIVLAFQKA 207
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 GSSPSVKCSPRGKNIQGGRTLSVPLEERQDSGTCTCTVSQDNVYEFKIDIVLAFQKA 180
QY 208 SSIYVKEGEQVEFSFPLAFVTEKLTGSGELVMQWQERASSSKSWITFDLKNKEVSKVKT 267
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 SSTYVKEGEQVEFSFPLAFVTEKLTGSGELRMQWQERASSSKSWITFDLKNKEVSKVKT 240
QY 268 QDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALAEKTKLHQEVNLVVMRATQLOKNT 327
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 QDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALAEKTKLHQEVNLVVMRATQLOKNT 300
QY 328 CEVWGPTSPKMLSLKLENKKAIVSKREKPVWVNLNPEAGMOCCLSDSGQVLLSNIKVL 387
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 CEVWGPTSPKMLSLKLENKKAIVSKQAKAVWVNLNPEAGMOCCLSDSGQVLLSNIKVL 360
QY 388 PTWSTPVP 396
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 PTWSTPVP 369

```

RESULT 8

```

ID 08H278 PRELIMINARY; PRT; 457 AA.
AC 08H278;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Lymphocyte antigen CD4.
OS Calitrich jaccus (Common marmoset).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callitrix.
OC NCBI_TaxID=9483;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22174698; PubMed=12186836;
RA Labonte J.A., Babcock G.J., Patel T., Sodroski J.;
RA "Blockade of HIV-1 infection of New World Monkey Cells Occurs
RT Primarily at the Stage of Virus Entry.";
RL J. Exp. Med. 196:431-445(2002).
DR EMBL; AF452616; AAN14532.1; -

```

DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0006955; P:immune response; IEA.
 DR InterPro; IPR000973; CD4_TcAg.
 DR InterPro; IPR003599; IG_Like.
 DR InterPro; IPR007110; IG_Like.
 DR InterPro; IPR003596; IG_v.
 DR Pfam; PF00047; Ig; 2.
 DR PRINTS; PR00692; CD4TCANTIGEN.
 DR SMART; SM00409; IG; 3.
 DR SMART; SM00406; IG; 1.
 DR PROSITE; PS50835; IG_Like; 1.
 SQ SEQUENCE 457 AA; 50878 MM; 718CFDB78D97F598 CRC64;

Query Match 47.2%; Score 1610.5; DB 6; Length 457;
 Best Local Similarity 78.8%; Pred. No. 3.4e-114;
 Matches 312; Conservative 35; Mismatches 48; Indels 1; Gaps 1;

QY 1 MNRGVPFPHLLVQLALLPAATQGNKVVLGKGDVTELTCTASOKSIOFHMKNSQIK 60
 DB 1 MNGGIPFPHLLVQLALLPAATQGNKVVLGKGDVTELTCTASOKSIOFHMKNSQIK 60
 QY 61 ILNGSGFLLTKGSPKLNDRASRLMDQGNPPLIKLKIEDSDTYICEVEDQKEVQL 120
 DB 61 ILGSGSVTVKGSGKLNDRIDSKSSWDGSPPLIRNVQVEDSTYICEVESKKEVEL 120
 QY 121 LVFGLTANSDTHLLQGSLTLTLSPGSSPVQCRSPRKNIOGKTLVSQLELDPSG 180
 DB 121 QVFGTLVNPDTHLIQGSLTLTLSPGSSPVQCRSPRKNIOGKTLVSQLELDPSG 180
 QY 181 TWCTVQLONQKKEFKIDIVLAFOKASSIYKKEGQVSPPLATVTEKLTGSGELMW 240
 DB 181 TWKCTVQHLLELV-FKINIVLAFOQASSIYKKEGQVSPPLATVTEKLTGSGELMW 239
 QY 241 QAERASSKSMITFDLKNKEVSVKRVTPQPKLQMGKKLPHLTLPOALPOYAGSGNTLLA 300
 DB 240 QAERASSKSMITFDLKNKEVSVKRVTPQPKLQMGKKLPHLTLPOALPOYAGSGNTLLA 299
 QY 301 LEAKTGKLGHEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKAVSKKEKPVW 360
 DB 300 LKGTGKLGHEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKAVSKKEKPVW 359
 QY 361 LNPEAGMOCLLSDSGVLLSNTIKVLTPTWSTVPEP 396
 DB 360 LNPEAGMOCLLSDSGVLLSNTIKVLTPTWSTVPEP 395

RESULT 9

Q8H2T7 PRELIMINARY; PRT; 457 AA.
 AC Q8H2T7;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Lymphocyte antigen CD4.
 OS Salimti scuturus (Common squirrel monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Salimti.
 OX NCBI_TaxID=9521;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22174698; PubMed=12186836;
 RA Labonte J.A., Babcock G.J., Patel T., Sodroski J.;
 RT "Blockade of HIV-1 infection of New World Monkey Cells Occurs
 RT Primarily at the Stage of Virus Entry.";
 RL J. Exp. Med. 196;431-445(2002).
 DR EMBL; AF452617; AAN14533.1; -
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0006955; P:immune response; IEA.
 DR InterPro; IPR000973; CD4_TcAg.
 DR InterPro; IPR003599; IG_Like.
 DR InterPro; IPR007110; IG_Like.
 DR InterPro; IPR003596; IG_v.
 DR Pfam; PF00047; Ig; 2.

DR PRINTS; PR00692; CD4TCANTIGEN.
 DR SMART; SM00409; IG; 3.
 DR SMART; SM00406; IG; 1.
 DR PROSITE; PS50835; IG_Like; 1.
 SQ SEQUENCE 457 AA; 50899 MM; B164DA89E70C575A CRC64;

Query Match 46.4%; Score 1584.5; DB 6; Length 457;
 Best Local Similarity 78.3%; Pred. No. 3.2e-112;
 Matches 310; Conservative 32; Mismatches 53; Indels 1; Gaps 1;

QY 1 MNRGVPFPHLLVQLALLPAATQGNKVVLGKGDVTELTCTASOKSIOFHMKNSQIK 60
 DB 1 MNGGIPFPHLLVQLALLPAATQGNKVVLGKGDVTELTCTASOKSIOFHMKNSQIK 60
 QY 61 ILNGSGFLLTKGSPKLNDRASRLMDQGNPPLIKLKIEDSDTYICEVEDQKEVQL 120
 DB 61 ILGQVNFVTVRGSKLTDRLDSKSSWDGSPPLIRNVQVEDSTYICEVESKKEVEL 120
 QY 121 LVFGLTANSDTHLLQGSLTLTLSPGSSPVQCRSPRKNIOGKTLVSQLELDPSG 180
 DB 121 QVFGTLVNPDTHLIQGSLTLTLSPGSSPVQCRSPRKNIOGKTLVSQLELDPSG 180
 QY 181 TWCTVQLONQKKEFKIDIVLAFOKASSIYKKEGQVSPPLATVTEKLTGSGELMW 240
 DB 181 TWKCTVQHLLELV-FEINIVLAFOQASSIYKKEGQVSPPLATVTEKLTGSGELMW 239
 QY 241 QAERASSKSMITFDLKNKEVSVKRVTPQPKLQMGKKLPHLTLPOALPOYAGSGNTLLA 300
 DB 240 QAERASSKSMITFDLKNKEVSVKRVTPQPKLQMGKKLPHLTLPOALPOYAGSGNTLLA 299
 QY 301 LEAKTGKLGHEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKAVSKKEKPVW 360
 DB 300 LKGTGKLGHEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKAVSKKEKPVW 359
 QY 361 LNPEAGMOCLLSDSGVLLSNTIKVLTPTWSTVPEP 396
 DB 360 LNPEAGMOCLLSDSGVLLSNTIKVLTPTWSTVPEP 395

RESULT 10

Q7Z5WL PRELIMINARY; PRT; 470 AA.
 AC Q7Z5WL;
 DT 01-OCT-2003 (TREMBLrel. 25, Created)
 DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Splice;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Bhat N.K.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.T., Wang J., Hsieh F.,
 RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Cantinici P., Prange C.,
 RA Raha S.S., Loughlino N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Harte S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzyzanski M.I., Skalska U., Smalins D.E., Schermer A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Spleen;
 RA Strausberg R.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC053984; AAH53984.1; -
 KW Hypothetical protein
 SQ SEQUENCE 470 AA; 51204 MW; 778CF34521483F1A CRC64;
 Query Match 37.4%; Score 1277; DB 4; Length 470;
 Best Local Similarity 48.1%; Pred. No. 8,8e-89;
 Matches 232; Conservative 33; Mismatches 104; Indels 178; Gaps 15;
 QY 30 LGKKDPTVELTCTAS--QKKSIQFHW-----KNSQIITLGNQGSFLLTKGSPK--LND 78
 DB 30 LVQPGGSLRLSCVYASQFTLNNDMHWVRQIGKLEWVSKIGTAGDRVYVAGSVKGRFTIS 89
 QY 79 RADSRSLWDQGNFPLIINKIKIEDSDTYICEVEDQKEVQLLVFGLTANSDTHILQGS 138
 DB 90 RENAKDSLYLQNM-----SLRVGDAAVYVC-----ARGGRMAPLGAFDIWGQ 133
 QY 139 LTLTLESPGSSPSVQCRSPRGKNIQSGKTLVSQLELQDSGTWTCTVLQNGKVEFKID 198
 DB 134 TWTVSSASTKGSPVFLPAPSSKSTSG--TALAGCL----- 168
 QY 199 IVVLAFOKASIVYKKEGQVEFSPFLAFTVEKLTGSGELMWQAEKASSKSWITFDLKN 258
 DB 169 -----VKDYPPEPYTVS-----NNSGALTSG----- 189
 QY 259 KEYSVAVRTQDPRLQWKKLPLHLTLPOALPOVAGSGNLTLALAEATGKHQEVNLVWR 318
 DB 190 -----VH-TPAVL-QSSGIYSLSSVTVVPSLSLTQTYI----- 222
 QY 319 ATOLQKILTCBVGFTSPKMLMLSLKLENKAKVSKKEKVVWVNLPEAGMOCILSDSGV 378
 DB 223 -----CNV-----NHRP----- 229
 QY 379 LIESNIKVLPTWSTPVEPKSCDKHTTCCPCPAPELLGGPSVFLPFPKPKDTLMISRTPEV 438
 DB 230 ---SNTKV---DKKVEPKSCDKHTTCCPCPAPELLGGPSVFLPFPKPKDTLMISRTPEV 282
 QY 439 TCVVVDVSHEDPEVKENMYVDGVEVHNATKPREEQYNSTRVSVLTVLHODMLNGKEY 498
 DB 283 TCVVVDVSHEDPEVKENMYVDGVEVHNATKPREEQYNSTRVSVLTVLHODMLNGKEY 342
 QY 499 KCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAV 558
 DB 343 KCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAV 402
 QY 559 EWESNQCPENNYKTTTPVLDSGSPFLYSKLTVDKSRWQGNVFCSVNHEALHNHYTQK 618
 DB 403 EWESNQCPENNYKTTTPVLDSGSPFLYSKLTVDKSRWQGNVFCSVNHEALHNHYTQK 462
 QY 619 SLTSLSPG 625
 DB 463 SLTSLSPG 469
 RESULT 11
 0727P5 PRELIMINARY; PRT; 469 AA.
 AC 0727P5;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]

RP SEQUENCE FROM N.A.
 RC TISSUE=Spleen;
 RA MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L.; Feingold E.A.; Grouse L.H.; Derge J.G.;
 RA Klausner R.D.; Collins F.S.; Wagner L.; Scheinen C.M.; Schuler G.D.;
 RA Altschul S.F.; Zeeberg B.; Buetow K.H.; Schaefer C.F.; Bhat N.K.;
 RA Hopkins R.F.; Jordan H.; Moore T.; Max S.I.; Wang J.; Hsieh F.;
 RA Datchenko L.; Marusina K.; Farmer A.A.; Rubin G.M.; Hong L.;
 RA Stapleton M.; Soares M.B.; Bonaldo M.F.; Casavant T.L.; Scheetz T.E.;
 RA Brownstein M.J.; Usdin T.B.; Toshiyuki S.; Carninci P.; Prange C.;
 RA Raha S.S.; Loquellano N.A.; Peters G.J.; Abramson R.D.; Mullaly S.J.;
 RA Bosak S.A.; McEwan P.J.; McKernan K.J.; Malek J.A.; Gunaratne P.H.;
 RA Richards S.; Morley K.C.; Hale S.; Garcia A.M.; Gay L.J.; Hulyk S.W.;
 RA Villalón D.K.; Muzny D.M.; Sodergren E.J.; Lu X.; Gibbs R.A.;
 RA Fahy J.; Helton E.; Kettelman M.; Madan A.; Rodriguez S.; Sanchez A.;
 RA Whiting M.; Madan A.; Young A.C.; Shevchenko Y.; Bouffard G.G.;
 RA Blakesley R.W.; Touchman J.W.; Green E.D.; Dickson M.C.;
 RA Rodriguez A.C.; Grimwood J.; Schmutz J.; Myers R.M.; Butlerfield Y.S.;
 RA Krzywinski M.I.; Skalska U.; Smalish D.E.; Scherch A.; Schein J.E.;
 RA Jones S.J.; Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Spleen;
 RA Strausberg R.;
 RL Submitted (Apr-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC051328; AAH51328.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 469 AA; 51395 MW; C8D5BE12BAAF795C CRC64;
 Query Match 37.1%; Score 1266; DB 4; Length 469;
 Best Local Similarity 96.7%; Pred. No. 6.1e-88;
 Matches 236; Conservative 0; Mismatches 4; Indels 4; Gaps 1;
 QY 382 SNIKVLPTWSTPVEPKSCDKHTTCCPCPAPELLGGPSVFLPFPKPKDTLMISRTPEV 441
 DB 229 SNTKV---DKKVEPKSCDKHTTCCPCPAPELLGGPSVFLPFPKPKDTLMISRTPEV 284
 QY 442 VVDVSHEDPEVKENMYVDGVEVHNATKPREEQYNSTRVSVLTVLHODMLNGKEY 501
 DB 285 VVDVSHEDPEVKENMYVDGVEVHNATKPREEQYNSTRVSVLTVLHODMLNGKEY 344
 QY 502 VSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAV 561
 DB 345 VSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAV 404
 QY 562 SNGQCPENNYKTTTPVLDSGSPFLYSKLTVDKSRWQGNVFCSVNHEALHNHYTQK 621
 DB 405 SNGQCPENNYKTTTPVLDSGSPFLYSKLTVDKSRWQGNVFCSVNHEALHNHYTQK 464
 QY 622 LSPG 625
 DB 465 LSPG 468
 RESULT 12
 08NF17 PRELIMINARY; PRT; 509 AA.
 AC 08NF17;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE FLJ00385 protein (Fragment).
 GN FLJ00385.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Spleen;

RA Jikuya H., Takano J., Kikuno R., Nagase T., Ohara O.;
RT "the nucleotide sequence of a long cDNA clone isolated from human
RT spleen."
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL, AK090464; BAC03445.1; -.
DR PIR, A45874; A45874.
DR InterPro; IPR007110; Ig-1-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig1.3.
DR SMART; SM00407; IGc1.3.
DR PROSITE; PS50835; IG_LIKE.3.
DR PROSITE; PS00290; IG_MHC; 2.
DR NON_TER
FT 1
SQ SEQUENCE 509 AA; 5611 MM; 0894980768863C CRC64;

Query Match 37.1%; Score 1265; DB 4; Length 509;
Best Local Similarity 91.7%; Pred. No. 8.1e-88;
Matches 231; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

QY 395 EPKSCDHTPCPCPAPAPLGGPSVFLPPPKDTLMISRTPEVTCVVDVSHDEPEVKF 454
DB 209 EPKSCTDTPPCPCPAPLGGPSVFLPPPKDTLMISRTPEVTCVVDVSHDEPEVKF 268
QY 455 NMVYDGEVHNNAKTPREEQYNSTYRVSVLTVLHQDLNGEKYCKVSNKALPAPIEKT 514
DB 269 NMVYDGEVHNNAKTPREEQYNSTYRVSVLTVLHQDLNGEKYCKVSNKALPAPIEKT 328
QY 515 ISKAKGPREPOVTLPPSRDELTKNOVSLTCLVKGFPSPDIAVEMENGGPENNYKTP 574
DB 329 ISKTKGQPREPOVTLPPSRDELTKNOVSLTCLVKGFPSPDIAVEMENGGPENNYKTP 388
QY 575 PVLDSGDSFFLYSKLTVDKSRMQGNVFSQVMEHALHNHYTQSLSPGLQDETCAE 634
DB 389 PVLDSGDSFFLYSKLTVDKSRMQGNVFSQVMEHALHNHYTQSLSPGLQDETCAE 448
QY 635 AODGELDLGWT 646
DB 449 AODGELDLGWT 460

RESULT 13
QY 72351 PRELIMINARY; PRT; 482 AA.
AC Q72351;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein DKFZp686N02209.
GN DKFZp686N02209.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Human rectum tumor;
RA Bleecker H., Boeher M., Mewes H.W., Weil B., Amid C., Oeanger A.,
RA Fobio G., Han M., Wiemann S.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BX58118; CAD98026.1; -.
KW Hypothetical protein.
SQ SEQUENCE 482 AA; 52852 MM; EDA75F1901D1A034 CRC64;

Query Match 37.0%; Score 1264.5; DB 4; Length 482;
Best Local Similarity 96.3%; Pred. No. 8.2e-88;
Matches 235; Conservative 0; Mismatches 5; Indels 4; Gaps 1;

QY 382 SNIKLPWSTPVEPKSCDHTPCPCPAPAPLGGPSVFLPPPKDTLMISRTPEVTCV 441
DB 242 SNTKV---DKVEPKSCDHTPCPCPAPAPLGGPSVFLPPPKDTLMISRTPEVTCV 297
QY 442 VVDVSHDEPEVKFMVYDGEVHNNAKTPREEQYNSTYRVSVLTVLHQDLNGEKYCK 501

DB 298 VVDVSHDEPEVKFMVYDGEVHNNAKTPREEQYNSTYRVSVLTVLHQDLNGEKYCK 357
QY 502 VSNKALPAPIEKTISKAKGPREPOVTLPPSRDELTKNOVSLTCLVKGFPSPDIAVEME 561
DB 358 VSNKALPAPIEKTISKAKGPREPOVTLPPSRDELTKNOVSLTCLVKGFPSPDIAVEME 417
QY 562 SNGPENNYKTPPVLDSGDSFFLYSKLTVDKSRMQGNVFSQVMEHALHNHYTQSL 621
DB 418 SNGPENNYKTPPVLDSGDSFFLYSKLTVDKSRMQGNVFSQVMEHALHNHYTQSL 477
QY 622 LSPG 625
DB 478 LSPG 481

RESULT 14
QY 096P08 PRELIMINARY; PRT; 679 AA.
AC Q96P08;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Factor VII active site mutant Immunofugate.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21477448; PubMed=11593034;
RA Hu Z., Garen A.;
RT "Targeting tissue factor on tumor vascular endothelial cells and tumor
RT cells for immunotherapy in mouse models of prostatic cancer."
RL Proc. Natl. Acad. Sci. U.S.A. 98:12180-12185(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Hu Z., Garen A.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF272774; AAK58686.2; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000152; Asx hydroxyl 5.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR00742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_11.
DR InterPro; IPR06209; EGF_11c.
DR InterPro; IPR002383; GLA_blood.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR007110; Ig-1-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001254; peptidase_S1.
DR InterPro; IPR003134; peptidase_S1A.
DR InterPro; IPR000294; vitk_dep_GLA.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00594; GLA; 1.
DR Pfam; PF00047; Ig1.2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00010; EGFBLD.
DR PRINTS; PR00001; GLABLOOD.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00069; GLA; 1.
DR SMART; SM00069; GLA; 1.
DR SMART; SM00020; TYP_SPC; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 1.

DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS00011; GLUT CARBOXYLATION; 1.
 DR PROSITE; PS50835; IG_Like; 2.
 DR PROSITE; PS00290; IG_MHC; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_SER; 1.
 DR PROSITE; PS00135; TRYPSIN_HIS; 1.
 SQ SEQUENCE 679 AA; 75552 MW; 0B0023AE70A067A1 CRC64;

Query Match 36.8%; Score 1258; DB 4; Length 679;
 Best Local Similarity 100.0%; Pred. No. 4.2e-87;
 Matches 211; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 395 EPKSCDKHTPCPCPAPELLGGPSVFLFPPPKDITMISRPEVTCVVVDVSHEDPEVKF 454
 DB 448 EPKSCDKHTPCPCPAPELLGGPSVFLFPPPKDITMISRPEVTCVVVDVSHEDPEVKF 507
 QY 455 NMVVDGVEVHNAKTKRREQGYNSTYRVSVLTITLVHODMNGKYEKCKVSNKALPAPIEK 514
 DB 508 NMVVDGVEVHNAKTKRREQGYNSTYRVSVLTITLVHODMNGKYEKCKVSNKALPAPIEK 567
 QY 515 ISKAGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTP 574
 DB 568 ISKAGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTP 627
 QY 575 PVLDSGSEFLYKSLTVDSKRWQGNVFCSSVMEHLNHYTKSLSPG 625
 DB 628 PVLDSGSEFLYKSLTVDSKRWQGNVFCSSVMEHLNHYTKSLSPG 678

RESULT 15

ID 09XS78 PRELIMINARY; PRT; 455 AA.
 AC 09XS78;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE T-cell surface glycoprotein CD4.
 OS Delphinapterus leucas (Beluga whale).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti;
 OC Monodontidae; Delphinapterus.
 NC NCBI_Taxid=9749;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Thymus;
 RA MEDLINE=99216435; PubMed=10199913;
 RA Romano T.A., Ridgway S.H., Felten D.L., Quaranta V.;
 RT "Molecular cloning and characterization of CD4 in an aquatic mammal,
 the white whale Delphinapterus leucas.";
 RL Immunogenetics 49:376-383(1999).
 RL EMBL; AF071799; AAD23738.1; -.
 DR HSSP; P01730; 1W1Q.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:006955; P:immune response; IEA.
 DR InterPro; IPR000973; CD4_TcRg.
 DR InterPro; IPR007110; IG_Like.
 DR InterPro; IPR003596; IG_v.
 DR Pfam; PF00047; Ig1; 3.
 DR PRINTS; PR00692; CD4TCANTIGEN.
 DR SMART; SM00406; IG1; 1.
 DR PROSITE; PS50835; IG_Like; 2.
 SQ SEQUENCE 455 AA; 50499 MW; AA532FD4411AA5D1 CRC64;

Query Match 34.7%; Score 1185.5; DB 6; Length 455;
 Best Local Similarity 60.2%; Pred. No. 7.8e-82;
 Matches 254; Conservative 46; Mismatches 107; Indels 15; Gaps 5;

QY 1 NMRGVPFRLLLVLTALIPATQGNKVVLGKGGDTVELTCTASQKKSIOFHKNQSNQIK 60
 DB 1 MDPRTSLRHLFLVLQVLMPLAGTQGGKVVILGKAGELAIELPKCAKSNKSLFPMKNSYQTK 60

QY 61 ILNGSEFLTKGSPSKLNRADSRRLMDQGNFPLIKKLIKEDSDTYICEVEDQKEEYOL 120
 DB 61 ILGRIGYFMHKGASLHRSVESKINLMDQGSFPLVTKOLEVDSCTYICEVEDKKEIYEL 120
 QY 121 LVFGLTANSDFHLLQGSLITLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDPSG 180
 DB 121 QVFRLLTASSDFRLLLGQSLITLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDPSG 180
 QY 181 TWTCVTLNQKKVEKIDIVLAFQKASSIYKKEGEQVEPFLAFYVEKLTGSGEL-W 239
 DB 181 TWTCVTLNQKKVEKIDIVLAFQKASSIYKKEGEQVEPFLAFYVEKLTGSGEL-W 239
 QY 240 WQAEASASSKSWITFDLKNKEYSVVRVTPDPLQMGKLLPLHLTLPQALPOYAGSGNLT 299
 DB 229 LQAKGNSSPESVITTKLNGKVTYVAKARDLKLRLSKALPLHLTLPQALPOYAGSGNLT 298
 QY 300 ALEATGKLHDEVNILVNRATQLOKRLTCEVNGPPTSPTKMLSLKLENKARKVSKKEPVW 359
 DB 299 NL-TKGKLYGEVNLVNRVTKSPNSLTCEVNGPPTSPTLISLKKENSMRVSDDQKLV 356
 QY 360 VLNPAAGMWQCLISDQVLESNIKVLPTWSTPYEPKSCDKHTPCPCPAPELLGGPSV 419
 DB 357 VLGPEAGMWQCLISDQGVLESKVKILP-----LAHAMPKLLAVLGGITSL 406
 QY 420 FL 421
 DB 407 LL 408

RESULT 16

ID 086TT2 PRELIMINARY; PRT; 354 AA.
 AC 086TT2;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Human full-length cDNA clone CS0D101YF20 of placenta of Homo sapiens
 DE (Human) (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NC NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Genoscope;
 RL Submitted (FEB-2003) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Li W.B., Gruber C., Jessee J., Polayes D.;
 RT "Full-length cDNA libraries and normalization.";
 RT Submitted (FEB-2003) to the EMBL/Genbank/DBJ databases.
 RL EMBL; BX248278; CAD62606.1; -.
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
 DR InterPro; IPR007110; IG_Like.
 DR InterPro; IPR003597; IG-cl.
 DR InterPro; IPR003006; IG_MHC.
 DR Pfam; PF00047; Ig1; 1.
 DR SMART; SM00407; IGcl; 3.
 DR PROSITE; PS50835; IG_Like; 3.
 DR PROSITE; PS00290; IG_MHC; 2.
 KM Plasmid.
 FT NON_TER.
 SQ SEQUENCE 354 AA; 39125 MW; 23B80BF4D2B87A92 CRC64;

Query Match 34.3%; Score 1171; DB 4; Length 354;
 Best Local Similarity 92.6%; Pred. No. 7e-81;
 Matches 214; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

QY 395 EPKSCDKHTPCPCPAPELLGGPSVFLFPPPKDITMISRPEVTCVVVDVSHEDPEVKF 454
 DB 123 EPKSCDKHTPCPCPAPELLGGPSVFLFPPPKDITMISRPEVTCVVVDVSHEDPEVKF 182

```
QY 455 NMVYDGVENHNAKTKPREEQNSTYRVSVLTVLHODWLNGKEYCKVSNKALPAPIEKT 514
D 183 KMVYDGVENHNAKTKPREEQNSTFRVSVLTVLHODWLNGKEYCKVSNKALPAPIEKT 242
QY 515 ISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVWESNGQPENNYKTP 574
D 243 ISKTKGQPREPOVYTLPPSRDEMTNQVSLTCLVKGFYPSDIAVWESSGQPENNYNTTP 302
QY 575 PVLSDSGSFLLYSKLTVDKSRMGOGNVSCVMHEALHNHTOKSLSLSPG 625
D 303 PMLSDSGSFLLYSKLTVDKSRMGOGNIPSCVMHEALHNHTOKSLSLSPG 353

RESULT 17
Q8N4Y9 PRELIMINARY; PRT; 521 AA.
AC Q8N4Y9
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Primary B-Cells;
RA Strauberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC033178; AAH33178.1; -.
DR PIR; A60764; A60764.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00407; IGL; 3.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
DR PROSITE; PS00290; IG_MHC; 2.
KM Hypothetical protein.
SQ SEQUENCE 521 AA; 57156 MW; 2AC7D22E72D6CAA2 CRC64;

Query Match 34.2%; Score 1167; DB 4; Length 521;
Best Local Similarity 92.2%; Pred. No. 2.4e-80;
Matches 213; Conservative 9; Mismatches 9; Indels 0; Gaps 0;

QY 395 EPKSCDKHTHTCPGPAPPELLGAPSVFLPPPKKDTLMISRTPEYTCVVVDVSHEDPEVKF 454
D 290 EPKSCDTPPCPCAPPELLGAPSVFLPPPKKDTLMISRTPEYTCVVVDVSHEDPEVQF 349
QY 455 NMVYDGVENHNAKTKPREEQNSTYRVSVLTVLHODWLNGKEYCKVSNKALPAPIEKT 514
D 350 KMVYDGVENHNAKTKPREEQNSTFRVSVLTVLHODWLNGKEYCKVSNKALPAPIEKT 409
QY 515 ISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVWESNGQPENNYKTP 574
D 410 ISKTKGQPREPOVYTLPPSRDEMTNQVSLTCLVKGFYPSDIAVWESSGQPENNYNTTP 469
QY 575 PVLSDSGSFLLYSKLTVDKSRMGOGNVSCVMHEALHNHTOKSLSLSPG 625
D 470 PMLSDSGSFLLYSKLTVDKSRMGOGNIPSCVMHEALHNHTOKSLSLSPG 520

RESULT 18
Q8TC63 PRELIMINARY; PRT; 473 AA.
AC Q8TC63
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
```

```
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Strauberg R.;
RL Submitted (MAR--2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC025985; AAH25985.1; -.
DR GO; GO:000507; F:copper ion binding; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR000923; BlueCu_1.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00196; COPPER_BLUE; 1.
DR PROSITE; PS00835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 3.
KM Hypothetical protein.
SQ SEQUENCE 473 AA; 51986 MW; E29920B09BA369F5 CRC64;

Query Match 33.7%; Score 1150; DB 4; Length 473;
Best Local Similarity 52.1%; Pred. No. 4.2e-79;
Matches 260; Conservative 43; Mismatches 101; Indels 95; Gaps 14;

QY 171 VSQLELDGSG-----TWCTTVLQNGKRVKFKIDIVLAFQKASSTVY-----KKE 215
D 25 LSRQLQDESGPGLKPEVTLSTICTVSGD-----SVASSSYMGWVRQPP 69
QY 216 GEQVEFSFLAFYVEKKTGSGELMWMQERASSSKSWITPPLKNEVVKVKTQPKLQMG 275
D 70 GKGLEWIGTINP-----SGNMYY-----SPSLRSRVTSADMSNSFYL-----KLD-- 111
QY 276 KKLPLHLTLPOALPOYAGSGLTLALBAKTKLHQEVNLVVMRATOLQKNLTCEWGPPT- 334
D 112 -----SVTAADTAVYYCAAGHLVWGFAHWG-----QCKLVSVSPASTK-----GPSV 154
QY 335 SPKIMLSLKLENKAKVSKREK-----PVNV-----LNPBAGMOCLLSDSGVTLLESN 383
D 155 FPLAPCSRSTSESTAALGCLVDFYFPEPVTVSMNSGALTGVHTFPVAVLQSSGLYSLS-SS 213
QY 384 IKVLPTSTVEVERKSCDKTH-----TCPPCAPPELLGAPSVFLPPPKP 426
D 214 VVTVPSSSLGCTKTYTCVNDHKPSNTKVDKRVESKYGPCCPAPBFLGAPSVFLPPPKP 273
QY 427 KDTLMISRTPEYTCVVVDVSHEDPEVKFNMVYDGVENHNAKTKPREEQNSTYRVSVLT 486
D 274 KDTLMISRTPEYTCVVVDVSHEDPEVQFNMVYDGVENHNAKTKPREEQNSTYRVSVLT 333
QY 487 VLHODWLNGKEYCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTC 546
D 334 VLHODWLNGKEYCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTC 393
QY 547 LVNGFYPSDIAVWESNGQPENNYKTPPVLSDSGSFLLYSKLTVDKSRMGOGNVSCSV 606
D 394 LVNGFYPSDIAVWESNGQPENNYKTPPVLSDSGSFLLYSRLTVDKSRMGOGNVSCSV 453
QY 607 MHEALHNHTOKSLSLSPG 625
D 454 MHEALHNHTOKSLSLSG 472

RESULT 19
P79355 PRELIMINARY; PRT; 474 AA.
AC P79355
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
```

DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
 DE CD4 antigen precursor.
 OS Felis silvestris catus (Cat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
 NCBI_TaxID=9685;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Miyazawa T.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92165316; PubMed=1537604;
 RA Noriame J., Miyazawa T., Kawaguchi Y., Tohya Y., Kai C., Mikami T.;
 RT "A CDNA encoding feline CD4 has a unique repeat sequence downstream of
 the V-like region."
 RL Immunology 75:74-79(1992).
 DR EMBL; AB000483; BAA19124.1; -.
 DR HSSP; P01730; 1MIQ.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:006955; P:immune response; IEA.
 DR InterPro; IPR000973; CD4_TcAg.
 DR InterPro; IPR007110; IG_1like.
 DR InterPro; IPR003596; IG_v.
 DR Pfam; PF00047; Ig_3.
 DR PRINTS; PR00692; CD4TCANTIGEN.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS50835; IG_LIKE; 2.
 DR SIGNAL.
 FT SIGNAL.
 FT CHAIN 1 26 POTENTIAL.
 SQ SEQUENCE 474 AA; 52243 MW; D946DD4BEAD00EC CRC64;
 Query Match 31.8%; Score 1087; DB 6; Length 474;
 Best Local Similarity 57.6%; Pred. No. 2.6e-74;
 Matches 239; Conservative 49; Mismatches 99; Indels 28; Gaps 6;
 QY 1 NNRGVPRRLILVLTALPAAT-OGNKVVLGKGGTVELTCTASOKKSIQPHMKNNOI 59
 DB 1 MNRGAVFRLLLVLTALVMLKAAPVPOGKEVVLGAGTALPCQASQKKTMTTMRLSGV 60
 QY 60 KILGNQGSFL-TKGSPLKNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEV 118
 DB 61 KILSGHSLCLTGSSKLTREFSKILMDQGSFLVIKSLVAVDSGITCEVENKKEV 120
 QY 119 QLLVFGLTANSD-----THLLQGSLTLLTSPGSSPSVQCSPPK 161
 DB 121 ELLVFGLTAKVDPSSGSGSSSSSTSTSTSYLLQGSLTLLTSPSSSNPSVQMKGPKN 180
 QY 162 NIQGGKITSVSGLELQDSTGWTCTVLQNKVYFKIDIVLAFQKASSIVYKKEGOVEF 221
 DB 181 SSGVSHSLSLQLEQSGTCTCTVSGSQKTVFNINILVLAFRVSNVYVAKKEGOVEF 240
 QY 222 SFPLAFVTEKLTGSGELMWQAERASSKSMITFDLKNKEVSVKRYTQDPKLGKPLH 281
 DB 241 SFPLAFEDBNLGN--LRKKAEGAPSSLLMISFTLKNQLSVKEDPYSKLOMNSLPR 298
 QY 282 LTLFQALPOYAGSGNLTALAEKTKLHQEVNLVVMRATQLOKNTICEYWGFTSPKML 341
 DB 289 FLLPVLNLSRYAGSNLTALVD--KGLQGEVLLVMRATVQSGNNLTCEVGLPTSPDLT 356
 QY 342 LKLEKKEAVSKREKRVVWLNPEAGMOCCLSDSGVLLSENIKYL-----TWS 391
 DB 357 LKLKGOAAKVSQKQKRVVEDAAGTWCCLSHKOKVLLASFAEVLPPVLTTRTWT 411

RESULT 20
 061396

ID 061396 PRELIMINARY; PRT; 457 AA.
 AC 061396;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, last annotation update)

DE T-cell differentiation antigen.
 GN CD4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8804159; PubMed=2823269;
 RA Gorman S.D., Tourvelille B., Parnes J.R.;
 RT "Structure of the mouse gene encoding CD4 and an unusual transcript in
 brain."
 RL Proc. Natl. Acad. Sci. U.S.A. 84:7644-7648(1987).
 DR EMBL; M17080; AAA37402.1; -.
 DR EMBL; M17076; AAA37402.1; JOINED.
 DR EMBL; M17077; AAA37402.1; JOINED.
 DR EMBL; M17078; AAA37402.1; JOINED.
 DR EMBL; M17079; AAA37402.1; JOINED.
 DR HSSP; P01730; 1MBR.
 DR MGD; MGI:88335; CD4.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:006955; P:immune response; IEA.
 DR InterPro; IPR000973; CD4_TcAg.
 DR InterPro; IPR007110; IG_1like.
 DR InterPro; IPR003596; IG_v.
 DR Pfam; PF00047; Ig_2.
 DR PRINTS; PR00692; CD4TCANTIGEN.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS50835; IG_LIKE; 1.
 SQ SEQUENCE 457 AA; 51368 MW; 24AB19EDA285B5D0 CRC64;
 Query Match 28.9%; Score 985; DB 11; Length 457;
 Best Local Similarity 53.6%; Pred. No. 1.5e-66;
 Matches 210; Conservative 62; Mismatches 110; Indels 10; Gaps 7;
 QY 1 NNRGVPRFH-LILVLTALPAATQGNKVVLGKGGTVELTCTASOKKSIQPHMKNNOI 59
 DB 1 MCRALSRLRLILLLQLSLAVQEKTLVKGESAEELCESQKKTITVFTWKFSDR 60
 QY 60 KILGNQGSFLTKG--PSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKE 116
 DB 61 KILGHHGKGVILRGSSPSQF-DREDSKGAWEKGSFPLIINKLKIEDSDTYICELENRKE 119
 QY 117 EVOLLVPELTANSDPHLQGSLLTLLTLES-PPGSSPSVQCSPPKKNIQGGKITSVSGLE 175
 DB 120 EVELMVFYVTPSPGSLLLQGGSLTLLTLDNSKVSNPTECHKKGKGVSGSKVLSMSNLR 179
 QY 176 LQDSTGWTCTVLQNKVYFKIDIVLAFQKASSIVYKKEGOVEFSPPLAFVTEKLTGS 235
 DB 180 VQDSFPMWCTYTLTDKKNMFGKMTLSVLGFQSTAITAYKSEBSAEFSPLNFAE--NGW 237
 QY 236 GELMWQAERASSKSMITFDLKNKEVSVKRYTQDPKLGKPLH 295
 DB 238 GELMWKAKDSFPQOWISFISIKNKEVSQKSTKDKLQKRTPLTKIPQVSIQFASG 297
 QY 296 NLTALAEKTKLHQEVNLVVMRATQLOKNTICEYWGFTSPKMLSLKLEKKEAVSKRE 355
 DB 298 NLTALVD--KGLTQGEVLLVMRATVQSGNNLTCEYWGFTSPKMLTLLQENQAEARVSEQ 355
 QY 356 KPVWVNLNPEAGMOCCLSDSGVLLSENIKYL 387
 DB 356 KVVGVVAPETGLMOCCLSEGDVKMDSRIQVL 387

RESULT 21
 055054

ID 055054 PRELIMINARY; PRT; 433 AA.
 AC 055054;
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
 DE T4 surface glycoprotein (Fragment).

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86097446; PubMed=3501122;
RA Madden P.J., Molineux S.M., Madden D.E., Zimmerman K.A., Godfrey M.,
RA Alt F.W., Chess L., Axel R.;
RT "Structure and expression of the human and mouse T4 genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:9155-9159(1987).
RN [2]
RP SEQUENCE FROM N.A.
RA Madden P.J., Molineux S.M., Madden D.E., Zimmerman K.A., Godfrey M.,
RA Alt F.W., Chess L., Axel R.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF045882; AAC01764.1; -.
DR HSSP; P01730; 1MR.
DR MGD; MGI:88335; Cd4.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig_Like.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4CANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
FT NON_TER
FT TER
FT SEQUENCE 433 AA; 48590 MW; AB19330750A8499A CRC64;
SO

Query Match 28.0%; Score 957; DB 11; Length 433;
Best Local Similarity 51.0%; Pred. No. 1,8e-64;
Matches 205; Conservative 61; Mismatches 114; Indels 22; Gaps 7;

QY 24 QGNKRVLGKGDYELTCTASQKSIQFMKNSQIKILGNQ--SFLTKG--PSKLNDR 80
DB 1 QGKTLVLGKEGSAELPCSSQKKTIVTWKFSDDRKILGQHGKVLIRGSPSQF--DRF 59
QY 81 DSRSLMNGQNPFLIKLKLTIEDSTTYICEVEDQKEVQLVFGITANSDFHLLQGGSLT 140
DB 60 DSKGAMKESGFPILINKLMKEDSQTYICELENKEEVELWVFATFSPGTSLLQGGSLT 119
QY 141 LTLRS-PPGSSPVSQCRPRGNKIQGGKTLVSQLELDQSGNWTCTVONOKVEFKDI 199
DB 120 LTLDSNKNVNPFLTECKKKGKGVSGKVLMSNLRVDSDFMNCVTLLDQKMFQKTL 179
QY 200 VLAIFQKASSIVYKKEGQVEFSPFLAFTVEKLTGSGELMWQAERASSSKMIFDLKNX 259
DB 180 SVLGQSFATITAYKSEGSAAEFSPFLNPAEE--NMGELMWKAERKDFQWISFSIKNX 237
QY 260 EVSVKRVYQDRLQMGKLLPHLTLPOLPYAGSGNLTLLAKTGLHDEVLVYVRA 319
DB 238 EVSVQKSTKDLTKLTKETLPLTKIPQVSLQFAGSGNLTLLD--KGTLHDEVLVYVRA 295
QY 320 TOLQNLTCCEWGPSPKLTMLSLKLENKAKVSKREKVVWYLVNPEAGMOCILDSGQVL 379
DB 296 AQLNNTLTCEWGPSPKRLTLTKOENKARVSEKQKVQVVAPEGTGLMOCLISBGDKVK 355
QY 380 LESNIKVLPTWSTVPEKSCDKTHTCPAPABELLGSPVFL 421
DB 356 MDSRIQVLSRGVNVFLAC-----VLGGSFATL 384

```

```

GN IGHCI.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RA Wagner B.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98383416; PubMed=9717671;
RA Wagner B., Overesch G., Sheoran A., Holmes M., Richards C.,
RA Leibold W., Radbruch A.;
RT "Organization of the equine immunoglobulin heavy chain constant region
genes. III. Alignment of c-mu, c-gamma, c-epsilon and c-alpha genes.";
RL Immunobiology 159:105-119(1998).
DR EMBL; AJ30675; CAC44624.1; -.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_2.
DR SMART; SM00407; IGV; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 2.
FT NON_TER
FT SEQUENCE 337 AA; 37438 MW; A60BF2B01DEFD1F6 CRC64;
SO

Query Match 26.9%; Score 919; DB 6; Length 337;
Best Local Similarity 61.3%; Pred. No. 1e-61;
Matches 173; Conservative 38; Mismatches 43; Indels 28; Gaps 5;

QY 371 LLSDDGVLLESNIKV-LPTWSTP-----VEP-----KSCDKTHTC 405
DB 56 VLQSSGYSLSSTWTVASTWTSETTYICNVHAASNKVDKRIERIDNNQKVDMS--KC 114
QY 406 PCPAPABELLGGPSVFPFPPKCDTLMISRTPEVTCVVDVSHEDPEKVMYVDGVEVHN 465
DB 115 PCPAPABELLGGPSVFPFPPKCDTLMITRPEVTCVVDVSGENPDVCFMVMYDGVVET 174
QY 466 AKTPREQVNSTYRVSVUTVLHOMLNKEKCKKSNKALPAPIKTSKAKGQREP 525
DB 175 ATTPKEQVNSTYRVSVUTVLHOMLNKEKCKKSNKALPAPIKTSKAKGQREP 234
QY 526 QVYTLPSRDELTFKQVSLCLVKGFPYSDIAVESNGQP--ENNYKTPPVLDSDGSF 583
DB 235 QVYTLAHPBELSKSKSVCLVKDFPFPININEMQNGPELETKSTTQAQDSGSGY 294
QY 584 FLVSKLTVDKSRWQGNVFSQVMEALNNHYTQKSLSPG 625
DB 295 FLVSKLSDVDRNRWQGTTFQGVVHEALNNHYTQKNSKNPG 336

```

```

RESULT 22
Q95M34 PRELIMINARY; PRT; 337 AA.
AC Q95M34;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Immunoglobulin gamma 1 heavy chain constant region (Fragment).

```

```

RESULT 23
Q7TMK1 PRELIMINARY; PRT; 470 AA.
AC Q7TMK1;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CZECH 11; TISSUE=Breast tumor;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Dege J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,

```

RA Diachenko L., Martusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldi M.F., Casavani T.L., Schetz T.E.,
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loughellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blackley R.W., Touchman J.W., Green E.D., Dickson J.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smallus D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences." ;
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CZECH II; TISSUE=Breast tumor;
 RA Struhsberg R.;
 RL SUBLMIT (AUG-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC055910; AAH55910.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 470 AA; 51727 MW; 6D9EADFD996BB090 CR664;

 Query Match 25.1%; Score 857; DB 11; Length 470;
 Best Local Similarity 39.1%; Pred. No. 8,6e-57;
 Matches 205; Conservative 58; Mismatches 129; Indels 132; Gaps 15,

 Oy PGSSPSVCGCRSRGNKIQ-----GKLSVSOLEDSGWTCVTVLONOKVRFK 196
 Db 33 PGASVXICKCA-SGYTFTGYMHWKQSHKLEMGIVNPENSGTGS---YNOK---FK 84

 Oy 197 IDIVLAAQKASSIYKKEGEQVERFPLAFVEKLTGSGELMWOAERASSKSWITFDL 256
 Db 85 -GKALTYLWDKSSSTAYME-----LNSLTSEDSAVYICATYYISGSITWYFDV 129

 Oy 257 --KNKEVSVKRVTQDPKLOMGKKLPYHLTLPGALPOYAGSG---NLTALAEK----- 304
 Db 130 WGAGTGVTVVSSAT-----TTAPSVYPLVPGCGDTSGSSVTLGLVKGYPPE 175

 Oy 305 -----TGKLGHEVNLV-----YMKATLOLQKLTCEWGPSTPKMLMS 341
 Db 176 PVTYKMYNGALSSGGRVTSVYLQSGFSLSLVTPSSITWPSQTIVICVNAHPAS----- 229

 Oy 342 LKLEKKAQKVSREKRPVWVLANPEAGMOCCLSDSGVLLSNIKYLPMTSPVPEKSCDK 401
 Db 230 -----KTELKRIIP-----RIKPSITP--PGS--- 250

 Oy 402 THTCPPCAPELLGSPSVFLPPPKKDTLMTSRPEVTCVVVDVSHEDPEVKFMNYVDGV 461
 Db 251 -----SCPNGNITGSPSVIFPPPKKDALMTISLTGKVTGVVDVSEDDPDVHVSFVDNK 305

 Oy 462 EVNNAKTPREQKQNSTRVVSVLTVLHODMLNGEYKCYKSNKLPAPIETKTSKAGQ 521
 Db 306 EVNHTAMTPREKQKQNSTRVVSALEPIQODMWKGEKFCCKVNNKLPAPIETKTSKPKGR 365

 Oy 522 PREPOVYTLPRPDELTKNOVSLTCLVGYGFPSDIAYVMESNGOEYNNYKTPTPVLDSG 581
 Db 366 AOTPOVYTIIPPREQKSKKQVSLTCLVNTNPFSEALISVMERNGELEDQYKTPTPVLDSG 425

 Oy 582 SFFLYSKLTVDKSRWQGNVSCSYVMHALLNHYTKSLSPG 625
 Db 426 TYFLYSKLTVDTPDSWLGEIFFCSSVYHEALNHNHQTKNLSLSPG 469

DE Gammal heavy chain of Mab7 (Fragment).
GN IGH-4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10990;
RN [1]
RP SEQUENCE FROM N.A.
RA Wilde K.G., Yu X., Ekramoddoullah A.K.M., Misra S.;
RT "Cloning of cDNA encoding for anti-white pine blister rust monoclonal
RT antibody (Mab 7, its light and heavy chains) and construction of a
RT single chain antibody (scFv).";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF152372; AAD40243.1; .
DR PIR; B45837; B45837.
DR PDB; 1COK; 11-SEP-99.
DR PDB; 1I91; 25-DEC-02.
DR PDB; 1KCU; 11-MAY-02.
DR MGI; MGI:96446; Igh-4.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00406; Igv 1.
DR PROSITE; PS50835; Ig_LIKE; 4.
DR PROSITE; PS00290; Ig_MHC; 1.
FT NON_TER 1 1
FT TER 437 437
SQ SEQUENCE 437 AA; 48142 MW; 5C3A7BB3E7D697C CRC64;

Query Match 24.9%; Score 850.5; DB 11; Length 437;
Best Local Similarity 33.4%; Pred. No. 2,4e-56;
Matches 202; Conservative 74; Mismatches 142; Indels 187; Gaps 15;

OY 30 LGKKGDVVELTCTAS--QKSIQPHMKNSNOIKILNQGSFLTGTGPKLNDRAADRSRLM 87
DB 10 LVKGGSGIKLSKSCASGFTSSYAMSWRQTPPEKLEWASFSGGIYYTDSVGRFTYI 69
OY 88 ---DQGNFPLIKLAKIEDSDPTYICEVDDQKEEVQLVFGLTANSDPHLQGSGLTLTLE 144
DB 70 KDKQRNLISLQMSLRSEDTMYTC-----AGDYSAWGPGLTVTVS 112
OY 145 SPSSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLQNKKEFKIDIVLAF 204
DB 113 AAKTTPSYVPLAB-----GSAQTNSMTL----- 138
OY 205 QKASIIYKKEGEVEYSEFPLAFVEKLTGSGELMQQAEARSSSKSMITPDLKKEVSVK 264
DB 139 -----GCLVKKYFPEPVT-----NNSGISLSSG----- 162
OY 265 RVTDPTKLQMGKKLPLHLTLPGQALPQ--YAGSGNLTLALEAKTGKLGQEVNLVYMRATQL 322
DB 163 -----VH-TTPAVLQSLYTLSSSVT-----VPSSTWP 189
OY 323 QKNLTCEVWGPTSPKMLSLKENKEAKVSRREKPVVNLNPEAGMOCLLSDSGVLLS 382
DB 190 SEVTTCVAVHPAS-----SKVKDK----- 209
OY 383 NIKVLPTWSTPVEPKSCDKTKHTCP--CPAPELLGSPVFLPPPKPKYDTLMISRTPEVTC 440
DB 210 -----IYPRDGG--CKPCTCTVPEV--SSVFPPKPDVLTITLTPEKVC 251
OY 441 VVVDVSHEDPEVKKMNVVDGVEVYNATKREBOYNTSYVSVLTVLHODMLNGKXYKC 500
DB 252 VVVDISKQDPEVQSMFVDVEVHTAQTQPREDFNSTFSSVSLPITMDQMLNGKKEKC 311
OY 501 KVSNAKLPAPIEKTIKSKAQGPREFQVYTLPPSRDELTLKQVSLTCLVKGFPSDIAVEW 560
DB 312 RVNSAAPAPAPIEKTIKTKGRPKAPQYVYIIPPRREGMAKQKVSITCMTITDFPEPDIWVW 371
OY 561 ESNQGPENNTKTPPVVLDSQGSFFLVSKLTVDKSRWQQGVNFGCSVNHGALAHNYTKSL 620
DB 372 QMNGQPEANNTKTPPVIMDTGSGYFVSVSKLVNQKSMWAGNTFTCSVLHDEGLAHNHTEKNL 431

OY 621 SLSPG 625
DB 432 SHSPG 436

RESULT 25

OBR3V9 PRELIMINARY; PRT; 469 AA.

ID OBR3V9
AC OBR3V9;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypoetical protein.
GN IGH-4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC024405; AAH24405.1; -.
DR PIR; B45837; B45837.
DR MGD; MGI:96446; Igh-4.
DR InterPro; IPR007110; Igh-4.
DR InterPro; IPR003006; Igh_MHC.
DR InterPro; IPR003596; Igh_v.
DR Pfam; PF00047; Igh_3.
DR SMART; SM00406; Igv_1.
DR PROSITE; PS50835; Igh_LIKE; 4.
DR PROSITE; PS00290; Igh_MHC; 1.
KW Hypoetical protein.
SQ SEQUENCE 469 AA; 51976 MW; 534793F155D05457 CRC64;

Query Match 24.9%; Score 850.5; DB 11; Length 469;
Best Local Similarity 34.9%; Pred. No. 2.7e-56;
Matches 189; Conservative 75; Mismatches 109; Indels 169; Gaps 15;

OY 147 PGSSPSVOC-----RSPRGK-----NIOGKTL5 170
DB 33 PGSLRLSCAASGFTDYWSWNPQPKALEWLGFRKANGYTTYASVSKGRFTIS 92
OY 171 -----VSOLELQDSGTCTVYQNGKVEFKIDIVLAFOKASIVYKKEGEQV 219
DB 93 RDNOSGSIYLOMNAARADSAIYCA---RDRRSY-----YSGTSFAYMGQGITLV 141
OY 220 EPS-----PFLAFTVEKLTGSGELMWQAEKASSSKSWITFDLKNKEVSVKRYTQDP 270
DB 142 TVSAKTTTPPSVYPLA-----PGS-----AAQNSMWLTGCLVKGY----- 177
OY 221 KLQMGKCLPLHLTPQALPOYAGSGNLTALAEAKTGKLGHOEV-----NLVYMRATQLOKN 325
DB 178 -----FPEPVTVTWNSGSISSGVHTPEPAVLQSDLYTLSSVTPVPSSTWPSQT 224
OY 326 LTCEVWGPSTPKMLSLKLENKEAVSKREKRVWVLANPEAGMWQCLSDSGVLLSNIK 385
DB 225 VTCVNAHNAS-----STKVDK----- 241
OY 386 VLPWSTVEPEKSCDKHTTCPP--CPABELGGSPVFLFPKPKDTLMISRTPEVTCVV 443
DB 242 -----IVPRDCG---CKPCICTVPEV---SSVFIFFPKPKDVLITLTLPKVTQVV 286
OY 444 DVSHEDPEVKFNWYDGEVYVNAKTKPREEQVNSTYRVSVLTVLHQMVLNGKEKCVS 503
DB 287 DISKDDEPQSFWDVDEVHTAQTKPREEQFNSTFRVSSELPIMHQMLNGKEKCRVN 346
OY 504 NKALPAPLEKTISSAKQPREPOVTLTPSRDELTKNOVSLTCLVKGFPYSDIAVWESN 563
DB 347 SAAPFAPELEKTISSKGRPKAPQVYTIIPPKQGMKDKVSLTCLMTDFPEDDIVEMQNM 406
OY 564 GQPENNYKTPPVLDSDGSFLLYKLTVDKSRMGGNVFSCSVMEALHNHYTKSLSL5 623

DB 407 GQPAENYKNTQPIMDTDGSYFYVYSKLVQKSNWEAGNTFTCSVLHEGLAHNHTEKSLSHS 466
OY 624 PG 625
DB 467 PG 468

RESULT 26

O99LC4 PRELIMINARY; PRT; 463 AA.

ID O99LC4
AC O99LC4;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Similar to RIKEN cDNA 1810060009 gene.
GN IGH-4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC003435; AAH03435.1; -.
DR PIR; B45837; B45837.
DR HSSP; P01842; 7PAB.
DR MGD; MGI:96446; Igh-4.
DR InterPro; IPR007110; Igh-4.
DR InterPro; IPR003006; Igh_MHC.
DR InterPro; IPR003596; Igh_v.
DR Pfam; PF00047; Igh_3.
DR SMART; SM00406; Igv_1.
DR PROSITE; PS50835; Igh_LIKE; 4.
DR PROSITE; PS00290; Igh_MHC; 1.
SQ SEQUENCE 463 AA; 51007 MW; EAA674C6BEC30783 CRC64;

Query Match 24.7%; Score 844; DB 11; Length 463;
Best Local Similarity 36.7%; Pred. No. 8.2e-56;
Matches 186; Conservative 72; Mismatches 120; Indels 132; Gaps 14;

OY 136 GQSLITLSPSPSSPSVQCRSPRGK-NIOGKTL5-----VSOLELQDSGTCTVYQNG 189
DB 61 GQGLEWGEIYIPSGGNTYSEKFKATLTDTSSSTAYVHMLSLTSEDSAAVFC----- 116
OY 190 QKKEVEFKIDIVLAFOKASIVYKKEGEQVEFS-----PFLAFTVEKLTGSGELMW 240
DB 117 -RSSYSYD-----PAYMGQTLVTVSAKTTTPSVYPLA-----PES----- 154
OY 241 QAERASSKSWITFDLKNKEVSVKRYTQDPKLGKCLPLHLTPQALPOYAGSGNLTALA 300
DB 155 -----AAQNSMWLTGCLVKGY-----FPEPVTVTWNSGSISSG 186
OY 301 LEAKTGKLGHOEV-----NLVYMRATQLOKNLTCEVWGPSTPKMLSLKLENKEAVSKRE 355
DB 189 VHTPEPAVLQSDLYTLSSSVTPSSTWPSSEYTCVNAHPAS-----STKVDK- 235
OY 356 KPVWVLANPEAGMWQCLSDSGVLLSNIKVLPTWSTVEPEKSCDKHTTCPP--CPABEL 413
DB 236 -----IVPRDCG---CKPCICTVPEV---SSVFIFFPKPKDVLITLTLPKVTQVV 253
OY 414 LGGSVPFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYDGEVYVNAKTKPREE 473
DB 254 -----SSVFIFFPKPKDVLITLTLPKVTQVVVDISKDDEPQSFWDVDEVHTAQTKPREE 310
OY 474 QNSTYRVSVLTVLHQMVLNGKEKCVSNKALPADIEKTISSAKQPREPOVTLTPS 533
DB 311 QFNSTFRVSSELPIMHQMLNGKEKCRVNSAAPPAPLEKTISSKGRPKAPQVYTIIPP 370
OY 534 RDELTKNOVSLTCLVKGFPYSDIAVWESNQGQPENNYKTPPVLDSDGSFLLYKLTVDK 593
DB 371 KEQMAKDKVSLTCLMTDFPEDDIVEMQNGQPAENYKNTQPIMDTDGSYFYISKLNVQK 430

Db 131 GTTITVSSAKTAPSVYELAP-----VCGDTT----- 157
Qy 198 DIVVLAFOKASSIVYKKEGOVEFSEPLAFVTEKLTGSGELMWOERASSKSMITFDLK 257
Db 158 -----GSSVTL--GCLVKGTFP-----EPVT-----LTNNGSLSSG----- 187
Qy 258 NKEVSVKRVTDOPKLOMGKKLPLHLTLPOALPO--YAGSGNLTALAEKTKLHOEVNLV 315
Db 188 -----VH-TFPAVLQSDLYTLSSSVT----- 207
Qy 316 VMARATOLQKULTCEVWGPTSPKLMSTLENKEAVNSKREKRWVNLNPEAGMOCCLSDS 375
Db 208 VTSTPSSQITCVNAHPAS-----STVDDKIEP-----R 238
Qy 376 GOVLLESNIKVLPTWSTVEPEKSCDKTHTCPP--CPAPELLGSPVFLFPKPKDTLMIS 433
Db 239 G-----PT-----IKP-----CPCKCPAPNLLGSPSVFLFPKPKIDVLMIS 275
Qy 434 RTPEVTCVVDVSHEDPEVKFNWYDGYEVHNAKTKPREQYNSTYRVSVLTVLHQDML 493
Db 276 LSPMVTGVVVDVSEDDPVOGISMFVNNVEVLTQAQTREEDYNSTLRVVSALPIQHODMM 335
Qy 494 NGKEYKCVSNKALPAPLEKITSKXGQPREPOVYTLPPSDELTKNQVSLTCLVKGTFP 553
Db 336 SGKEKCKCVNNKALPAPLEKITSKXGQVRAPOVYVLPPEPEEMTKKQVTLTCVTDMP 395
Qy 554 SDIAYEMESNGOPENNYKTPPVLDDSGFPFLYSKLTVDKSRMOQGNVSCSVNHEALHN 613
Db 396 EDIYEMTNNGKTELNTKTEPVLDDSGSYFMYSKLRVEKNWERNYSYSCSVNHEGLHN 455
Qy 614 HYTKSLSLSPG 625
Db 456 HHTKTSFRTPG 467

RESULT 29

Q99L25 PRELIMINARY; PRT; 473 AA.
ID 099L25;
AC 099L25;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to RIKEN cDNA 1810060009 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC003888; AA03888.1; -
DR HSSP; P01842; 7FAB.
DR InterPro; IPR007110; Ig_1like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
SQ SEQUENCE 473 AA; 52449 MW; BE9889B7986DA155 CRC64;

Query Match 23.7%; Score 810.5; DB 11; Length 473;

Best Local Similarity 47.4%; Pred. No. 3e-53; Indels 59; Gaps 7;

Matches 165; Conservative 47; Mismatches 77;

Qy 285 PQALPOYAGSGNLTALAEKTKLHOEV-----NLVNRATOLQKULTCEVWGPTSPKLM 339
Db 177 PEPVTLTNNGSLSGVHTFPAVLQSDLYTLSSSVTVSSTPSSQITCVNAHPAS----- 232
Qy 340 LSLKLENKAYSKREKRWVNLNPEAGMOCCLSDSGOVLLESNIKVLPTWSTVEPEKSC 399
Db 233 -----STVDDKIEP-----RG-----PT-----IKP----- 249

Qy 400 DKHTTCTP--CPAPELLGSPVFLFPKPKDTLMISRTPEVTCVVDVSHEDPEVKFNWY 457
Db 250 -----CPCKCPAPNLLGSPSVFLFPKPKIDVLMISLSPMVTGVVVDVSEDDPVOGISMF 304
Qy 458 VDGVEVHNAKTKPREQYNSTYRVSVLTVLHQDMLNGKEYKCVSNKALPAPLEKITSK 517
Db 305 VNNVEVLTQAQTREEDYNSTLRVVSALPIQHODMNGKEFKCKVNNKALPAPLEKITSK 364
Qy 518 AKGQPREPOVYTLPPSDELTKNQVSLTCLVKGTFSPSDIAYEMESNGOPENNYKTPPVL 577
Db 365 PKGSVRAPOVYVLPPEPEEMTKKQVTLTCVTDMPEDIVETNNGKTELNTKTEPVL 424
Qy 578 DSDGFPFLYSKLTVDKSRMOQGNVSCSVNHEALHNHYTKSLSLSPG 625
Db 425 DSDGSYFMYSKLRVEKNWERNYSYSCSVNHEGLHNHHTKTSFRTPG 472

RESULT 30

Q91Z05 PRELIMINARY; PRT; 473 AA.
ID 091Z05;
AC 091Z05;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN AU044919.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC010327; AA010327.1; -
DR MGD; MGI:2144967; AU044919.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR000345; Cytc_heme_BS.
DR InterPro; IPR007110; Ig_1like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00190; CYTOCHROME C; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
DR Hypothetical protein.
SQ SEQUENCE 473 AA; 51946 MW; CF625F008932AF12 CRC64;

Query Match 23.4%; Score 800.5; DB 11; Length 473;

Best Local Similarity 33.5%; Pred. No. 1.7e-52; Indels 179; Gaps 15;

Matches 204; Conservative 69; Mismatches 157;

Qy 30 LGKKGDTVELTCTAS--QKKSIOFHWKNSNOIKILG-----NQGSP-----LTGSPSKL 76
Db 30 LVKPGGSRKLSCAASGTFSDYGMHWRAQAPBKGLAVAINSGSTTIYADYVKGKRTI 89
Qy 77 NDRADSRSLMDQGNFPLIITKLTIEDSDTYICEVEDQKEBVOALLVGLTANSDTLLQG 136
Db 90 S-RDNARKTLF-----LQMTSLRSEDTAMYYCARLEMLRRIID-----YWG 128
Qy 137 QSLTLTESPPGSPSVQCRSPKGNIQGKTLVSQLELDSDSTWTCTVLQNKVBEK 196
Db 129 QGTTITVSSAKTTPPSVPLAPGCGDTTG----- 157
Qy 197 IDIVLAFOKASSIVYKKEGOVEFSEPLAFVTEKLTGSGELMWOERASSKSMITFDL 256
Db 158 -----SSVTL--GCLVKGTFPESVTV-----WNGSLSS----- 186
Qy 257 KNKEVSVKRVTDOPKLOMGKKLPLHLTLPOALPOYAGSGNLTALAEKTKLHOEVNLV 316

```

Db      187 -----VH-TFPALL-----QSGLYTMSSSVTV 207
Qy      317 MEATOLQKLTCEVWGPTSPKMLSLKLENKAKVSKREKPVWVLNPEAGMQLSDSG 376
Db      208 PSTSPSCVITCSVAHPAS-STTVDKLE-----PSCPISTINP----- 245
Qy      377 QVLESNIKVLPTWSTPVEPKSCDKTHTCPCPAPELLGSPVFLFPKPKDTLMISRP 436
Db      246 -----CPCKECHKCPAPLLEGSPVFLFPPIIKVLMISLTP 283
Qy      437 EYTCVYVVDVSHDEPEKFWYDGVENNAKTKPREBOYNSTRYVSVLTJLHQMUNGK 496
Db      284 KYTCVYVVDVSEDDPDVQISWFMVNVVHTAQTOTREDYNSTIRVVSALPIQHODMMSGK 343
Qy      497 EKCKVSNKALPAPIEKTSKAKGPREPOVYTLPPSRDELTKQVSLTCLVKGFPYPSDI 556
Db      344 EKCKVSNKALPSPERITSKITGLVAPQVYILPPAPQLSRKQVSLTCLVGNFPGI 403
Qy      557 AVEWESNGOPENNYKTPPVLDSDGSPFLYSKLTVDKSRMQOQNVFSCVMHEALHNHYT 616
Db      404 SVEWTSNGHTENYKDTAPVLDSDGYFYISKLDIKTSKCKTDFSCVNRHEGLKNYVL 463
Qy      617 QKSLSLSPG 625
Db      464 KKTISRSPG 472

RESULT 31
ID      Q8R3H6 PRELIMINARY; PRT; 474 AA.
AC      Q8R3H6;
DT      01-JUN-2002 (Tremblrel. 21, Created)
DT      01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT      01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE      Hypothetical protein.
GN      AU044919.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxId=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Strausberg R.;
RL      Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
DR      EMBL; BC025447; AAH25447.1; -.
DR      MGD; MG2:214467; AU044919.
DR      GO; GO:0005489; Electon transporter activity; IEA.
DR      GO; GO:0006118; P:electon transport; IEA.
DR      InterPro; IPR000345; Cytochrome_B5.
DR      InterPro; IPR007110; Ig-like.
DR      InterPro; IPR003006; Ig_MHC.
DR      InterPro; IPR003596; Ig_v.
DR      Pfam; PF00047; Ig_3.
DR      SMART; SM00406; IGV; 1.
DR      PROSITE; PS00190; CYTOCHROME_C; 1.
DR      PROSITE; PS50835; IG LIKE; 4.
DR      PROSITE; PS00290; IG_MHC; 1.
KM      Hypothetical protein.
SQ      SEQUENCE 474 AA; 51748 MW; 8608B576CD2874A CRC64;

Query Match      23.4%; Score 799.5; DB 11; Length 474;
Best Local Similarity 33.2%; Pred. No. 2,1e-52;
Matches 209; Conservative 71; Mismatches 155; Indels 195; Gaps 18;

```

```

Db      121 -----YG-----DYPDWKQGAIVTVSSAKTTPSPSYPLAPGCGDPTG----- 158
Qy      176 LODSGTWTCTVLQNKKEFKIDIVLAFQCAASSIVYKKEBOVFSPPLAFYHEKLTGS 235
Db      159 -----SSVTL-----GCLVVGYPPEVTVT----- 178
Qy      236 GELMWQARASSSKSWITFDLKNKEVSVKVTVQDPKLGKKLPLHLTLPOALPYAGSG 295
Db      179 -----MNSGLSS-----VH-TFPALL----- 195
Qy      236 NLTALAEKTKLHGEVNLVVMRATOLQKLTCEVWGPTSPKMLSLKLENKAKVSKRE 355
Db      196 -----QSGLYTMSSSVTVPSSTWSPQYTCVAHPAS-STTVDKLE-----PS 238
Qy      356 KPVWVLNPEAGMWOQLSDSQVLIESNIKVLPTWSTPVEPKSCDKTHTCPCPAPELLG 415
Db      239 GPISTINP-----CPCKECHKCPAPLLEG 263
Qy      416 GPSVFLFPKPKDTLMISRTPEVTCVYVVDVSHDEPEKFWYDGVENNAKTKPREBOY 475
Db      264 GPSVFLFPKPKDTLMISRTPEVTCVYVVDVSEDDPDVQISWFMVNVVHTAQTOTREDY 323
Qy      476 NSTYRVSVLTJLHQMUNGKEXYCKVSNKALPAPIEKTSKAKGPREPOVYTLPPSRD 535
Db      324 NSTIRVVSALPIQHODMMSGKEPKCKVNNKDLSPERITSKITGLVAPQVYILPPAP 383
Qy      536 ELTKQVSLTCLVKGFPYSDIAVEWESNGOPENNYKTPPVLDSDGSPFLYSKLTVDKSR 595
Db      384 QLSRDVSLTCLVGVGFNPGDISVWTSNGHTENYKDTAPVLDSDGYFYISKLDIKTSK 443
Qy      596 WQOQNVFSCVMHEALHNHYTOKSLSLSPG 625
Db      444 WEKTDSPSCVNRHEGLKNYILKTIISRSPG 473

RESULT 32
ID      Q96BY0 PRELIMINARY; PRT; 613 AA.
AC      Q96BY0;
DT      01-DEC-2001 (Tremblrel. 19, Created)
DT      01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT      01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE      Hypothetical protein.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxId=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Strausberg R.;
RL      Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
DR      EMBL; BC011857; AAH11857.1; -.
DR      PIR; S15590; S15590.
DR      InterPro; IPR007110; Ig-like.
DR      InterPro; IPR003006; Ig_MHC.
DR      InterPro; IPR003596; Ig_v.
DR      Pfam; PF00047; Ig_5.
DR      SMART; SM00406; IGV; 1.
DR      PROSITE; PS50835; IG LIKE; 5.
DR      PROSITE; PS00290; IG_MHC; 3.
KM      Hypothetical protein.
SQ      SEQUENCE 613 AA; 67273 MW; 31214203FB8421E7 CRC64;

Query Match      12.2%; Score 417.5; DB 4; Length 613;
Best Local Similarity 22.2%; Pred. No. 3.7e-23;
Matches 151; Conservative 117; Mismatches 280; Indels 131; Gaps 24;

```

```

Qy      8 RHLLVLTQALLP-----AATGKKNVVLGKKGDTVELTCTASQKSIQFHWKN-- 61
Db      2 KHLWFFLLVLAAPRWVLSQVQLQSGGPGLVKPSFTLSLTCTVSGSGSISSYVWMIQBPAG 61

```

```

Oy 62 LGND--GSFLTKPSEKLNDRADSRSL--WDQGNFPIIKLKIEDSDTYICEVEDQKE 116
Db 62 KGLEWIGRIYTSSTNYPKSLKRVMTMSVDTSKQFSLKLSVAADTAAYVC---ASQ 117
Oy 117 EVOLLVFLVLTANSPTDHLIQGSLTITLESPPGSSPS---VQCSSPRKAIQGGKTLSSV 172
Db 118 PWEIPTVGL-----FYWGQDTLVTVSSGASATLTPPLVSC----- 153
Oy 173 QLELDSSGTWTCTVLQNOKKEKVIDIVLAFQ--KASSIVYKKEQGEVFSFPLAFTVEK 231
Db 154 ----ENSPSDTSSVAVGLADFLPDSITTFSMKTKNNNDISTGQ-----FPSVLKGGK 203
Oy 232 LTGSGELMWQABRASSSSKSWITFDLKXKXSVKRVTDQPKLQMGKQLPLHL--TLPOALP 289
Db 204 YAATSQVLLPESKDV-----MQGDEHVCKVQHPNKNKEKNVLPVIAELPKVVS 253
Oy 290 QYA-----GSGNITLLEAKTGLHQBVLVVMR-----ATQIQKRLTCEWNGPT 334
Db 254 VFVPRDGFENPRKSKLIQATGFSPPRIQVSWLREGKQVSGVTTDQVQAEKESGPT 313
Oy 335 SPKMLSLKLENKEKAVSKREKPVVVLNPEAGMQLSDSGQVLE--SNIKVLPTWSTP 393
Db 314 TYKTTSTLTIKESD-----WL---SOSMFTCRVDHRLGLPQOMASSMCPDQDTA 360
Oy 394 VEPKSCDKTHTCPCPAPABELLGGPSVFLFPKPKDTLMISRTPEVTCVVDVSHEDPEVK 453
Db 361 IR-----VFALPPS--FASIFLTKSTKLTCLVTLDTLYD--SVT 395
Oy 454 FNVVVDGEVNNATKPREBOYNSTYRVSVLTVLHQMVLNGKXKCVSKALPAPLEK 513
Db 396 ISWRQNGEAVKTHNTNISESHPNATFSAVGASLCEDMNSGERFTCTVHTDLPSPKQ 455
Oy 514 TISKAKGP--REPOVYTLPPSRDEL--KNQVSLTCLVKGFPSPDIIVMESNGQP--ENN 569
Db 456 TISPKGVALRPPVYLLPAPBOULNRESATITCLTVGFSPADVPVQMGKGPLSEK 515
Oy 570 YKTPPVLD--SDGSFFLYSKLTVDKSRWQGNVFCSSVMEHALNNHYTKSLSPGLQ 627
Db 516 YVTSFMPPEPQAPGRYFAHSILITVSEBEMNGETVTVVAHEALPNRTERTVDSK---- 571
Oy 628 LDFTCAEQDDELGLWTT 646
Db 572 -TEGEVSADEEGFENLWAT 589

RESULT 33
O7TMT6 PRELIMINARY; PRT; 614 AA.
AC O7TMT6;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6NCR; TISSUE=Hematopoietic Stem Cell;
RX MEDLINE=22386257; PubMed=12477932;
RA Strauberg R., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.B., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Muliyil J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,

```

```

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzyzanski M.I., Skalka U., Smalhus D.E., Schenck A., Schein J.E.,
RA Jones S.J., Maira M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6NCR; TISSUE=Hematopoietic Stem Cell;
RA Strauberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC053409; AAI53409.1; -.
KW Hypothetical protein.
SQ SEQUENCE 614 AA; 67746 MW; 839BAF3B8D124F89 CRC64;

Query Match 12.24; Score 416; DB 11; Length 614;
Best Local Similarity 22.74; Pred. No. 4.8e-23;
Matches 156; Conservative 101; Mismatches 233; Indels 198; Gaps 27;

Oy 30 LGKKGDTVELTCTASQKSIQFHMKN-----SNOIKILGN-----QGSFLTKGP 73
Db 30 LVKPGASVKISKAS--GYAFSSSMVMWVKQPKGKLEWIGRIYVPGDGTNYNGK--KGG 86
Oy 74 SKLN--DRADSRSLMDQGNFPIIKLKIEDSDTYICEVEDQKEVQLVFGLTANSDT 132
Db 87 ATLTDKSSSTAMQ-----LSLTSEDSAVYFC--ANDYSSYFAFWG----- 129
Oy 133 LLOGQSITLTLESP--PGSSPSVQGNRP--RGNKIQGGKTLSSVQLELDSDGTCTVLQ 189
Db 130 --QGLTVFAESQSFNVPFVLVSCSPSLSDKLVLMGCLARDPFLSTISFTNN----- 181
Oy 190 QKXVEFIDIVLAFQKASSIVYKKEGGEVFSFPLAFTVEKLTGSGELMWQABRASSK 249
Db 182 -----YQNNAEVIQ-----GIR 193
Oy 250 SWITFDLKXKXSVKRVTDQPKLQ-----GKKLPLHLTLPOALPOVAGSG 295
Db 194 TFPPLRTGKGLVLTNSQLSPKSLIBSGDEYLWCKIHGKKNDDLVHP-- 243
Oy 296 NLTLALAKTKGLHQBVLV-----MRATQIQKRLTCEWNGPTSPKMLSLKENKEA 349
Db 244 -----AVAEKNPVNVVFPVRDGFSGPAPRKSXKLIQCATVFTPRPTVSVMLKQGLV 295
Oy 350 KVSREKPVVVLN-----PEA-----GMNQCLSDSGQVLESNIKVL 388
Db 296 ESGFTTDPVTIEKSGSTPQYKVIISTLTISEIDMLNLNVTCRVDHRLGTLFLK----- 348
Oy 389 TWSTPVEPKSCDKTHTCPCPAPABELLGGPSVFLFPKPKDTLMISRTPEVTCVVDVSH 448
Db 349 -----ANSTCTAASPSDIL-----TFIIPSPAD--IFLSISANITCLVSNLATY 392
Oy 449 DPEVKENVYDGEVNNATKPREBOYNSTYRVSVLTVLHQMVLNGKXKCVSKALP 508
Db 393 E-TLNTISSAQSEPELETKIKIMESHPNGTFFSAKGVASVCEBDMNRKKEVCTVTRDLP 451
Oy 509 APIEKTISKAKGPPE-----POVYTLPPSRDEL--KNQVSLTCLVKGFPSPDIIVMES 562
Db 452 SPKKKFTSK---PNEVHKHPPAVYLLPAPBOULNRESATVTVCLVKGFSFPAISVQWLO 507
Oy 563 NQO--PENNYKTPPVLD--SDGSFFLYSKLTVDKSRWQGNVFCSSVMEHALNNHYTK 618
Db 508 RQGLLQDEKAVTSAPRPPAPRPFYTHSLITVTEEBMSNGERTVTVSHALPHLVTER 567
Oy 619 SLSLSPGLQDFTCAEQDDELGLWTT 646
Db 568 TVDKS-----TEGEVVAEEEGFENLWTT 590

RESULT 34
O8WUK1 PRELIMINARY; PRT; 613 AA.

```



```

OX NCBI_TaxID=95568;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.E., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RL mangabeys (primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR EMBL: AF057387; AAC25131.1; -.
DR HSSP: P01730; 1CDY.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR InterPro: IPR007110; Ig-like.
DR Immunoglobulin domain; T-cell; MHC; Transmembrane.
KV Immunoglobulin domain; T-cell; MHC; Transmembrane.
FT DOMAIN 1 19 IG-LIKE V-TYPE DOMAIN.
FT NON_TER 1 1
FT DOMAIN 20 >86 IG-LIKE C2-TYPE DOMAIN.
FT DISULFID 49 78 BY SIMILARITY.
FT NON_TER 86 86
SQ SEQUENCE 86 AA; 9406 MW; 2BD97A9E819582AB CRC64;

Query Match 11.6%; Score 400; DB 6; Length 86;
Best Local Similarity 89.4%; Pred. No. 4.8e-23;
Matches 76; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 107 YICEVEDQKEEVQLVFGLTANSPTHLQGSLLTLESPPGSSPSVQCSPRKNIOGG 166
DB 1 YICEVEDKKEVELLVFGLTANSPTHLQGSLLTLESPPGSSPSVQCSPRKNIOGG 60
OY 167 KTLVSQLELDSDGTWCTVLOK 191
DB 61 RTLSPOLERQDSGTWCTVSDOK 85

RESULT 37
OY 077594 PRELIMINARY; PRT; 86 AA.
ID 077594
AC 077594;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
DE (Fragment).
OS Cercopithecus mitis.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheidae; Cercopithecus.
OX NCBI_TaxID=36225;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.E., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RL mangabeys (primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR EMBL: AF057381; AAC25125.1; -.
DR HSSP: P01730; 1CDY.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR InterPro: IPR007110; Ig-like.
DR Immunoglobulin domain; T-cell; MHC; Transmembrane.
KV Immunoglobulin domain; T-cell; MHC; Transmembrane.

```

```

FT NON_TER 1 1
FT DOMAIN <1 19 IG-LIKE V-TYPE DOMAIN.
FT DISULFID 20 >86 IG-LIKE C2-TYPE DOMAIN.
FT DISULFID 49 78 BY SIMILARITY.
FT NON_TER 86 86
SQ SEQUENCE 86 AA; 9417 MW; 371CA39EF58182AB CRC64;

Query Match 11.6%; Score 397; DB 6; Length 86;
Best Local Similarity 89.4%; Pred. No. 8.2e-23;
Matches 76; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 107 YICEVEDQKEEVQLVFGLTANSPTHLQGSLLTLESPPGSSPSVQCSPRKNIOGG 166
DB 1 YICEVEDKKEVELLVFGLTANSPTHLQGSLLTLESPPGSSPSVQCSPRKNIOGG 60
OY 167 KTLVSQLELDSDGTWCTVLOK 191
DB 61 RTLSPOLERQDSGTWCTVSDOK 85

RESULT 38
OY 077599 PRELIMINARY; PRT; 86 AA.
ID 077599
AC 077599;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
DE (Fragment).
OS Theropithecus gelada (Gelada baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheidae; Theropithecus.
OX NCBI_TaxID=9565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.E., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RL mangabeys (primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR EMBL: AF057389; AAC25133.1; -.
DR HSSP: P01730; 1CDY.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR InterPro: IPR007110; Ig-like.
KV Immunoglobulin domain; T-cell; MHC; Transmembrane.
FT NON_TER 1 1
FT DOMAIN <1 19 IG-LIKE V-TYPE DOMAIN.
FT DISULFID 20 >86 IG-LIKE C2-TYPE DOMAIN.
FT DISULFID 49 78 BY SIMILARITY.
FT NON_TER 86 86
SQ SEQUENCE 86 AA; 9420 MW; 2BD97A98574EE9AB CRC64;

Query Match 11.6%; Score 397; DB 6; Length 86;
Best Local Similarity 88.2%; Pred. No. 8.2e-23;
Matches 75; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

OY 107 YICEVEDQKEEVQLVFGLTANSPTHLQGSLLTLESPPGSSPSVQCSPRKNIOGG 166
DB 1 YICEVEDKKEVELLVFGLTANSPTHLQGSLLTLESPPGSSPSVQCSPRKNIOGG 60
OY 167 KTLVSQLELDSDGTWCTVLOK 191
DB 61 RTLSPOLERQDSGTWCTVSDOK 85

```

```

RESULT 39
O8VCX7      PRELIMINARY;      PRT;      613 AA.
AC O8VCX7;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN IGH-6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Salivary gland;
RA Strubberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC018315; AAH18315.1; -.
DR MGD; MGI:96448; Igh-6.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; IgV_5.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PSS00835; IG_LIKE; 5.
DR PROSITE; PSS00290; IG_MHC; 3.
KM Hypothetical protein.
SQ SEQUENCE 613 AA; 67855 MW; 41A9384D4C22862 CRC64;

Query Match      11.6%; Score 396.5; DB 11; Length 613;
Best Local Similarity 22.7%; Pred. No. 1.5e-21;
Matches 157; Conservative 100; Mismatches 227; Indels 209; Gaps 27;

OY 30 LKKKQDYTELTCTAS-----QKSAIQFMKNSNOIKILNGGS--FLTKGPS 74
DB 30 LMKPAPASVISCATGYTFSSYWIEMVKORPGHGLEWIG----EILPGSGTNYNEKFKG 85
OY 75 KLNDSAD-SRSLMOGNFPLIIKLIKIDSDTYICEVDOKEVQL-----VFGTLA 127
DB 86 KATFTADTSSNTAYWQ-----LSSLTSDSAVYYC-----ARLGRWYFDVWG-- 128
OY 128 NSDTHLQOGSLTLTLESP--PGSSPSVOCRSP-RGKNIOGGKTLVSQLELQDSGTWTC 184
DB 129 -----AGTYTVSSESQSPFNVPFLVSCSPSLDKULVAMGCLARDLPSLTSTWN- 180
OY 185 TVLQNOQKVEFKIDIVLAFQKASSIVYKKEGOVEFSPFLAFTVEKLTGSGELWQAE 244
DB 181 -----YQNTVEVIO----- 189
OY 245 ASSSKSWITFDLAKKEVSKRYTQDPKLOM-----GKKLPHLTLTFOALPO 290
DB 190 -GITFTPLRTGKGLATSOVLSPKSLIBGSDEVLCKIYGGGNRLHVPIP----- 242
OY 291 YAGSGNLTLALBAKTKLHQEVLNV-----MRAIQLOKNTLCEVWGPSTPKMLSLKL 344
DB 243 -----AAAKNPANVNVFPFPGDSSGAPRKSKLICATNTNTPRPITVSMUK 289
OY 345 ENKEAKVSKREKRPVWLVN---PEA-----GMQCLSDSGOVLLEBN 383
DB 290 DQKLVEGSGFTTDPVTIENKSGTFQTYKIVLSTLISIDWLNANVTTCRDHNGFLFK-- 347
OY 364 IKVLVTSVPVPEPKSCDKHTKTPCPAPAPELLGSPSVFLFPPPKKOTLMSRTPEVTCVV 443
DB 348 -----NVSSTCAASPSDIL---TFIIPSPFAD-IFLSKSNLTLCLVS 386
OY 444 DVSHEDPEVKFMVYDGVGVHNAKTRPREQYNSYTRVVSVLTVLHODMLNKEYKCKYS 503
DB 387 NLATYE-TLNIWASQSGEPLETKIKIMESHNGFISAGVASCVEEDNNKKEVCTVT 445
OY 504 NKALPAPIEKTSKAKGQPRE-----POVYTLPPSRDELTL-KNOVSLTCLVGVGFPSDIA 557
DB 446 HDDLSPQPKFKISK-----PNEVHKIPBAVYLLPPAREQNLRESATVTCVGVGFPSADIS 501

```

```

OY 558 VEMESNGQ--PENNYKTTTPVLD--SDGSFLYSKLTVDKSRMOQGNVPSGVMEALHN 613
DB 502 VQMLRGQLLPQEKTYTSPMPPEGPAPGFYFTHSLITVTEEBMNGSETTYTCVGHALPH 561
OY 614 HYTKSLSPQLQDETCAEAKODELDGLMTT 646
DB 562 LVTERTVKSS-----TEGEVNAEEGPFENIMWT 589

RESULT 40
O77595      PRELIMINARY;      PRT;      86 AA.
ID O77595;
AC O77595;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3) (Fragment).
OS Cercopithecus galierlus chrysogaester.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Cercopithecus.
OX NCBI_TaxID=75569;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98120644; PubMed=9656488;
RX Harris E.E., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the mangabeys (primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
CC -I- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY SIMILARITY).
CC -I- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR EMBL; AF057382; AAC25126.1; -.
DR HSSP; P01730; ICXY.
DR GO; GO:0016021; C:Integral to membrane; IEA.
DR InterPro; IPR007110; Ig-1-like.
KM Immunoglobulin domain; Glycoprotein; T-cell; MHC; Transmembrane.
FT NON TER 1
FT DOMAIN 1
FT <1 19 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 20 >86 IG-LIKE C2-TYPE DOMAIN.
FT CARBOHYD 79 76 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 49 78 BY SIMILARITY.
FT NON TER 86
SQ SEQUENCE 86 AA; 9419 MW; A9D97A9E19582BE CRC64;

Query Match      11.6%; Score 395; DB 6; Length 86;
Best Local Similarity 88.2%; Pred. No. 1.2e-22;
Matches 75; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 107 YICEVEDQKEVQLVFGLTANSPTHLQOGSLTITLSPGSSPSVOCRRPRKNIOGG 166
DB 1 YICEVEDKKEVELLVFGLTANSPTHLLEGSLTITLSPGSSPSVOCRRPRKNIOGG 60
OY 167 KTLVSQLELQDSGTWTCVLANOK 191
DB 61 RTLSPQLERQDSGTWTCVNSODK 85

RESULT 41
O96BB9      PRELIMINARY;      PRT;      597 AA.
ID O96BB9;
AC O96BB9;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=B-cell;
 RA Strauberg R.;
 RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC015760; AAH15760.1; -
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003596; Ig_V.
 DR Pfam; PF00047; Ig; 5.
 DR SMART; SMO0406; IG; 1.
 DR PROSITE; PS00835; IG LIKE; 5.
 DR PROSITE; PS00290; IG_MHC; 3.
 KW Hypothetical protein.
 SQ SEQUENCE 597 AA; 65039 MW; 4FC3AD8ECE263D9 CRC64;

Query Match 11.5%; Score 392.5; DB 4; Length 597;
 Best Local Similarity 22.4%; Pred. No. 2.9e-21;
 Matches 143; Conservative 116; Mismatches 246; Indels 131; Gaps 24;

QY 30 LGKGGDTVELTCTAS--QKSIQPHKNSNQK-----ILNGSGFL---TKGPKL 76
 DB 30 LVQPGSGIRLSCASGFSFSSYAMWVQAPGKLEWYSAISGSGSYTYADSVKGRFTI 89
 QY 77 NDRADSRSLMDQGFLLIKNLKEDSDTYICEVEDQKEVQLVPLGLTANSPTHLLOG 136
 DB 90 S-RDSSRRLTYLQNN-----SLRAEDTAIVYCAKDPG-----YASGNYREDYWG 135
 QY 137 QSLTLTLSPGSSPS---VQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNNQK 192
 DB 136 QGLTWTSSGSAATLPLVSC-----ENSPSDISSAVAGCLA 174
 QY 193 VEFKTDIVLAFQ-KASSIVKKEGQVEFSPPLAFYTEKLTGSGELMWQERASSSKSW 251
 DB 175 QDFLPDSITFSWKYKNSDISSTRG-----FPSVLRGKYAASQVLLPSKDV----- 222
 QY 252 ITFDLKNKEVSKRYTODPKLQMGKLLPLHL--TLQALPOYA-----GSGNLTAL 302
 DB 223 ---WQGDENHYVCVQHNGKNGKKNVPLPVIAELPPKVSFVPRDPFGPKSKLIC 278
 QY 303 AKTGKHOENVLVWR-----ATOLQNLTCVWGPSTPKMLSLKENKEAKVSKR 354
 DB 279 QATGSPRQIQVSMRKGQVGGVTTDQVQAEKESGPTTYKVTSLTIKESD----- 332
 QY 355 EKPWVVLNPEAGMOCCLISDSQVLE-SNFKVLPWSTPVEPKSCDKTHTCPPCAP 413
 DB 333 ---WL---SOSMPTCRVDHRLGTFQONASSMCVDDQDAIR----- 367
 QY 414 LGGPEVFLPPPKOTLMISRTPEVTCVAVDVSHDEPKFMWYDVGVVNAKTKPRE 473
 DB 368 ---VFAPIPS-FASIFLTSTKLTCLVTDLTVD-SVTSWTRQNGEAVKTHNSES 420
 QY 474 QYNSTRVAVSVLTVLHODMLNGEKYCKVSNKALPAPIEKTISKAGOP-REPOVYTL 532
 DB 421 HPNATFSAVGSASICEBDMNGSERFTCTVYTHDLPSPKQITSRKQVLAHRPDVYLLP 480
 QY 533 SRDELT-KNQS/LTLVKGFPSPDIAMVESNGOP--ENNYKTTTPVLD--SDGSFFLYS 587
 DB 481 AREQLNLESATITCLVTFGSPADVFVQMQRGQPLSEPKYVTSAPMPEQAPGRYFAHS 540
 QY 588 KLTVDKSRMOQGNVSGVMHEALHNHYTOKSLSPG 625
 DB 541 ILTVSEEMTGETITCVVAHEALPNRYTERKVDKSTG 578

RESULT 42
 077598 PRELIMINARY; PRT; 86 AA.
 AC 077598;
 DT 01-NOV-1998 (Tremblrel. 08, Created)

DT 01-NOV-1998 (Tremblrel. 08, last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
 DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
 DE (Fragment).
 OS Papio sp. (baboon).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
 OC Cercopitheciinae; Papio.
 OC NCBI_TaxID=61183;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=98320644; Pubmed=9656488;
 RX Harris E.E., Disotell T.R.;
 RT "Nuclear gene trees and the phylogenetic relationships of the
 mangabeys (primates: Papionini).";
 RL Mol. Biol. Evol. 15:892-900(1998).
 CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
 RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
 SIMILARITY).
 CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
 ONE V-LIKE AND 1 C2-LIKE DOMAIN.
 DR EMBL; AF057388; AAC25132.1; -.
 DR HSSP; P01730; ICDY.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR InterPro; IPR007110; Ig-like.
 KW Immunoglobulin domain; Glycoprotein; T-cell; MHC; Transmembrane.
 FT NON TER 1
 FT DOMAIN 1
 FT DOMAIN 20
 FT CARBOHYD 79
 FT DISULFID 49
 FT NON TER 86
 FT 86
 SQ SEQUENCE 86 AA; 9433 MW; A9D97A98574E83BE CRC64;

Query Match 11.5%; Score 392; DB 6; Length 86;
 Best Local Similarity 87.1%; Pred. No. 2e-22;
 Matches 74; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 107 YICEVEQKEVQLVGLTANSPTHLLOGSLTLTLSPGSSPVQCRSPRGKNIQGG 166
 DB 1 YICEVEDKEKEVELLVGLTANSPTHLLEGSLTLTLSPGTSVSKCRSPRGKNIQGG 60
 QY 167 KTLVSQLELDGSGTWTCTVLQNNQK 191
 DB 61 KTLVSQLELDGSGTWTCTVNSQDOK 85

RESULT 43
 096GAG
 ID 096GAG PRELIMINARY; PRT; 614 AA.
 AC 096GAG;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
 DE Hypothetical protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=B-cell;
 RA Strauberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC009851; AAH09851.1; -.
 DR PIR; S15590; S15590.
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0003700; C:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR000005; HTHAAC.
 DR InterPro; IPR007110; Ig-like.


```

DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PR00047; IgV_5.
DR SMART; SM00406; IgV_1.
DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; 1.
DR PROSITE; PS00835; IG_LIKE; 5.
DR PROSITE; PS00290; IG_MHC; 3.
DR Hypothetical protein
KW Hypothetical protein
SQ SEQUENCE 614 AA; 67921 MW; 55EF536E77AA9BB CRC64;

Query Match
Beet Local Similarity 11.4%; Score 389; DB 4; Length 614;
Matches 150; Conservative 116; Mismatches 272; Indels 144; Gaps 26;

QY 11 LVLQALLPAATQGNKVVLG---KKGTVELTCTASQKSI--QFHKNSNQKIL-- 62
DB 7 ILFLVAAATDAVSQMLVQSGAEVKKTTGSSVSKVSCASGYTTRYRLHVRQAPQALQEW 66
QY 63 -----GNQGSFLTKGSPSKLNDRAISRSLMDQNFLLITKNLKIEDSDTYICEVEDQ 114
DB 67 MGMTPFNGN-----TNYAQKQODRVTTRDR-SMNTAMELSLSLSEDTAMTYC----- 115
QY 115 KEEVQLVLEGLTAN-SDTHLQGSGLTLTLESPPGSSPS-----VQCRSPRGKNIQSGKTL 169
DB 116 -----ARGYSSMNDADFIDWGQGTMTVSSGSAAPFLPLVSC----- 154
QY 170 SVSOLEQDSGTWCTVVLONQKKVEPKIDIVLAFO-KASIVYKKEGQVEFSPPLAT 228
DB 155 -----ENSPSDTSSVAAGCLADFLPDSITFSWKKNNSISSITRG-----FPSVLAR 201
QY 229 VEKLTGSGELMMQAEASASSKSMITFDLKNKEVSVRVTPQDLQMGKKLPLHL--TLPQ 286
DB 202 GGYAATSOVLLPSKQV-----MGCTEHHVCKQHPHNGKKEKVPPLVLAELRP 251
QY 287 ALPOVA-----GSGNLTLEAKTKGKHQEVNLVNR-----ATOLQKNLTCEVW 331
DB 252 KVSIVFVRPRDFPGNPRKSKLTCQATGSPRQIQVSWLAEGQVSGVTTDQVQAEAKS 311
QY 332 GFTSPFLMSTLKENKEAVSKREKPVVYLNPEAGMOCGLSDSGQVLE-SNIKVLPTW 390
DB 312 GFTYKVNSTLTIKESD-----WL---SOSMFTCRVDHRIGLTFQOANASSMCPVQ 358
QY 391 STPVEPKSCDKHTKTPRCAPALLGGPSVFLRPKPKDTLMSRPPEVNCVVVDVSHEDP 450
DB 359 DPAIR-----VFAIRPS-FAISFLTKSKLTCVLTDLTTYD- 393
QY 451 EVKFMVYDGVENVNAKTPREBOYNSTRVVSVLTLVHQDLNGEKVKYSNKAIPAP 510
DB 394 SVTISGTRNGEAVKTHNTISHSRPATISAVGEASICEDDNNSGERFTCTYTHDLPSP 453
QY 511 IKTTSKAKGQF-REPOVYTLTPPSRDELTKQVSLTCLVKGPRPSDIANVESNGQF-- 566
DB 454 LKQTSIRPRGVALHRRDVTLLPPARQLMLRESATITCLVTGFSPADVFNQMGQGPUS 513
QY 567 ENNYKTPPVLD--SPGSPFLVSKLTVDSKRNQGVFSCVWHEHLHHYTKOKSLSP 624
DB 514 PKRYVTSAPMEPEQAGRYFAHSILTVSEEWNTGTYTCVAHEALPRTVETKDKS- 572
QY 625 GLQDDECAEQDGGELDGLMTT 646
DB 573 ----TEGEVSADBEQGFENIMAT 590

RESULT 44
077601 PRELIMINARY; PRT; 86 AA.
AC 077601;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3) (Fragment).
OS Lophocoebus albigena albigena.

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Lophocoebus.
OK NCBI_TaxID=75568;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=96320644; PubMed=9656488;
RA Harris E.E., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RT mangabeys (primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR EMBL; AF057391; AAC25135.1; -.
DR GO; GO:0016021; C:integral to membrane, IEA.
DR InterPro; IPR007110; I9-1-like.
KW Immunoglobulin domain; T-cell; MHC; Transmembrane.
FT NON TER 1 1
FT DOMAIN <1 19 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 20 >86 IG-LIKE C2-TYPE DOMAIN.
FT DISULFID 49 78 BY SIMILARITY.
FT NON TER 86 86
SQ SEQUENCE 86 AA; 9463 MW; 2BD97A88464FE9AB CRC64;

Query Match
Beet Local Similarity 11.4%; Score 388; DB 6; Length 86;
Matches 74; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 107 YICEVEDQKEEVQLVLEGLTANSDTHLQGSGLTLTLESPPGSSPSVQCRSPRGKNIQSG 166
DB 1 YICEVEDQKEEVQLVLEGLTANSDTHLQGSGLTLTLESPPGSSPSVQCRSPRGKNIQSG 60
QY 167 KTLVSOLEQDSGTWCTVVLONQK 191
DB 61 RTLVSQLEKQDSGTWCTVVSQDK 85

RESULT 45
077600 PRELIMINARY; PRT; 86 AA.
AC 077600;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3) (Fragment).
OS Lophocoebus aetartimus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Lophocoebus.
OK NCBI_TaxID=75566;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=96320644; PubMed=9656488;
RA Harris E.E., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RT mangabeys (primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR EMBL; AF057390; AAC25134.1; -.
DR HSRP; P01730; 1CDY.

```

```
DR GO; GO:0016021; C:Integral to membrane; IEA.
DR InterPro; IPR007110; Ig-like.
KW Immunoglobulin domain; Glycoprotein; T-cell; MHC; Transmembrane.
FT NON_TER 1
FT DOMAIN 1
FT 1 19 IG-LIKE V-TYPE DOMAIN.
FT 20 >86 IG-LIKE C2-TYPE DOMAIN.
FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 49 78 BY SIMILARITY.
FT NON_TER 86
SQ SEQUENCE 86 AA; 9476 MW; A9D97A88464FE9BE CRC64;

Query Match 11.2%; Score 383; DB 6; Length 86;
Best Local Similarity 85.9%; Pred. No. 9.5e-22;
Matches 73; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

Qy 107 YICEVEDKEEVOLVFGLTANSPTHTLQGSLTLTLESPGSSPVGCRSPKRNIOGG 166
Db 1 YICEVEDKEEVOLVFGLTANSPTHTLQGSLTLTLESPGSSPVGCRSPKRNIOVG 60

Qy 167 KTLVSQLELDQSGTWTCTVLQNR 191
Db 61 RTLSTVPLERQDSGTWTCTVNSQDQK 85

RESULT 46
Q96AA6 PRELIMINARY; PRT; 618 AA.
AC Q96AA6;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Lymph;
RA Strauberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC017356; AAH17356.1; -.
DR PIR; S15590; S15590.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_5.
DR SMART; SM00406; IgV_1.
DR PROSITE; PS00835; IG_LIKE; 5.
DR PROSITE; PS00290; IG_MHC; 3.
KW Hypothetical protein.
SQ SEQUENCE 618 AA; 67758 MW; 96DBD4C7C69650A6 CRC64;

Query Match 11.2%; Score 383; DB 4; Length 618;
Best Local Similarity 21.8%; Pred. No. 1.6e-20;
Matches 149; Conservative 117; Mismatches 284; Indels 132; Gaps 25;

Qy 8 RHLLVLQALLP-----AATGKRVVLGKKGTVELTCTASQKSIQFHWKNSNQIKI 61
Db 2 KHLWFFLLVAAPRWVLSQVLQWAGLLKPSRTLSLTGCVVGGSPSGYVMSWIRDPG 61

Qy 62 LGNQ--GSFLTKGPKSKNDRAADRSRL--WDQGNPPLIKNLKIEDSDT--YICEVEDQKE 116
Db 62 KGLEWIGINHSSTNYPNLSKSRVTISVDTSKQLSLKSSVNAADTAVYYC----- 114

Qy 117 EVQLVLEFGLTANSDTHL---LQGSLLTTLSPGSSPS---VQCRSPKRNIOGGKTL 169
Db 115 --AAVITAPSPGTGRYGMADWVGQGTIVTVSSGSAAPTLPPLVSC----- 158

Qy 170 SVSQLELDQSGTWTCTVLQNRKVEFKIDIVLAFO--KASSIVYKKEGEQVEFSPPLAFT 228
Db 159 -----ENSPSDTSSVAVGCLAQDFLPSITFSWKYKNSDISSTRG-----FPSVLR 205
```

```
Qy 229 VEKLTSGELMWOAERASSSSKSWITFDLKNKEVSVKRVTDPPKIQWKGKPLPHL--TLPQ 286
Db 206 GGAATASQVLYLPSKDV-----MQGTDEHVCKVQHPNGNKEKNPLPVIAELP 255

Qy 287 ALQVYA-----GSGNLTALAEKTKGKLGHEVNLVNR-----ATQLQKLTGEVW 331
Db 256 KVSFVPPRDGFGNPNPKSKLICOATGFSPPQLOVSWLRGKOVGSGVTTDQVQAKES 315

Qy 332 GPTSPKMLSLKLENKEAKVSKREKPVWVLPNPEAGMOCLLSDSGVYLE-SNIKVLPTW 390
Db 316 GPTTKVTSLTLTIKESD-----KL---SQSMFTCVDRHGLTFQGNASMCVPDQ 362

Qy 391 STPEPKSCDKHTCCPCPAPELLGSPVFLPPPKPDTLMIKSTPEVTCVVDVSHEDP 450
Db 363 DTAIR-----VPAIDPS--FASIFLTKSTLTCLVTLDTLTYD- 397

Qy 451 EYKFNWYVDGEVHNNAKTRREQYNSTYVSVLTVLHODMLNGKRYCKVSKALPAP 510
Db 398 SVTISWTRNGEAVKTHNTNISESHPNATFSAVGBASICEDDMNSGERFTCTVHTDLPSP 457

Qy 511 IEKTSKAKGQP-REPQVYTLPPSRDEL-T-KNQVSLTCLVKGFPSPDIAYMESNGQP-- 566
Db 458 LKQTSRPRKVALHRPQVYLLPPAREQLNRESATITCLVTGSPBAPVQMGQGPLS 517

Qy 567 ENNYKTPVLD--SDGSFFLYSKLYDKSRMOQGNVSCSWHEALHNHYTKSLSLSP 624
Db 518 PEKVYTSAPWPEQAPGRYPFAHSILTVSEEMWTGETYTCVVAHEALPNHVTERTVKS- 576

Qy 625 GLQUDTCARQDGEIDGLWTT 646
Db 577 -----TSEGVSADEGEFENLMAT 594
```

```
RESULT 47
Q9BUI0 PRELIMINARY; PRT; 597 AA.
AC Q9BUI0;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Lymph;
RA Strauberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC002963; AAH02963.1; -.
DR HSSP; P01825; 7FAB.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_5.
DR SMART; SM00406; IgV_1.
DR PROSITE; PS00835; IG_LIKE; 5.
DR PROSITE; PS00290; IG_MHC; 3.
KW Hypothetical protein.
SQ SEQUENCE 597 AA; 65274 MW; 2DAFA8FB7E055861 CRC64;

Query Match 11.1%; Score 378.5; DB 4; Length 597;
Best Local Similarity 21.9%; Pred. No. 3.3e-20;
Matches 145; Conservative 115; Mismatches 274; Indels 127; Gaps 24;

Qy 8 RHLLVLQALLP-----AATGKRVVLGKKGTVELTCTASQKSIQFHWKNSNQIKI 61
Db 2 KHLWFFLLVAAPRWVLSQVLQWAGLLKPSRTLSLTGCVVGGSPSGYVMSWIRDPG 61

Qy 62 LGNQ--GSFLTKGPKSKNDRAADRSRL--WDQGNPPLIKNLKIEDSDT--YICEVEDQKE 116
Db 62 KGLEWIGINHSSTNYPNLSKSRVTISVDTSKQLSLKSSVNAADTAVYYC----- 114
```

```

Qy 117 EVQLLVFGITLANSSTHL---LOGQSLTLTLSPSSPS---VQCRSPRGKNIQGGKTL 169
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 115 --ARVITRASPCTDGRGMDVWGQGTITVTVSSGSAAPLFLPVLC----- 158
Qy 170 SVSQELDQSGTWTCTVLOKQKVEFKIDIVLAQ--KASSTVYKKEGQVFPPLAFT 228
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 159 -----ENSPSDTSSVAVGCLAQDFLPDSITFSMKYKNNSDISSTRG-----FPSVLR 205
Qy 229 VEKLTSGEELMMQAEASSSKSWITFDLNKKEVSVKRYTQDPKLGKKLPLHL--TLPQ 286
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 206 GGTAAATGQVLLPSKQV-----MGTDEHVCKVQHPNGKKEKNVPLPVIAELRP 255
Qy 287 ALPOYA-----GSGNLTALAEKTKLHGEVNLVNR-----ATOLQKNIACEVW 331
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 256 KVSVPFPRDGFPGNPKRSKLIQATGSPRQIQVSWLRBKGQVSGVTTDQVQAIAKES 315
Qy 332 GPTSPKMLSLKLENKEAKVSKREKPVVNLNPEAGMOCCLSDSGVILE--SNIKVLPW 390
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 316 GPTTKVSTLTITLKESD-----WL---SQSMFCRVDRHRLTFQONASSMCVDPQ 362
Qy 391 STPVPKSCDKHTKPCPCAPELLGSPVFLPRPKKOTLMISRTPEVTCVVVDVSHEDP 450
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 363 DTAIR-----VFALPPS--FASIFLTSTKTLCLVTLDTLTYD-- 397
Qy 451 EVKFMVYDQVEVHNAKTKPREEQYNSTYRVVSVLTVLIHQDMLNGEKYCKVSNKALPAP 510
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 398 STLTSTWNGEAVKTHNTNISHPATSAVGEASICEEDMNSGERFTCTVTHDLPSP 457
Qy 511 IEKTSKAKGQF--REPQVYTLPPSRDELTKNOVSLTCLVKGFPSPDIADVEMESNGQF-- 566
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 458 LKQITSRPKGVALHNPVDVLLPAPAEQLMLRESAITTCLVGFSPADVFMQMGQQLPS 517
Qy 567 ENNYKTPPVLD--SDGSFPLYSKLTVDKSRMOQGNVSCVMHEALNHTYQKSLSLSP 624
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 518 PEKVTSAPEPQAPGRYFAHSILTVEEEMNTGETYTCVVAHEALPNRYERTVDKST 577
Qy 625 G 625
      |
Db 578 G 578

RESULT 48
Q9BOB8 PRELIMINARY: PRT: 597 AA.
ID Q9BOB8:
AC 09BOB8:
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle, and Lymph.
RA Strausberg R.;
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC006180; AAH06180.1; -.
DR EMBL; BC001872; AAH01872.1; -.
DR HSSP; P01825; 7FAB.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG_5.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG_LIKE; 5.
DR PROSITE; PS00290; IG_MHC; 3.
KW Hypothetical protein.
SQ SEQUENCE 597 AA; 65300 MW; 2DAFAD50A6375851 CRC64;
Query Match 11.0%; Score 376.5; DB 4; Length 597;

```

```

Best Local Similarity 21.9%; Pred. No. 4.7e-20;
Matches 145; Conservative 115; Mismatches 274; Indels 127; Gaps 24;

Qy 8 RELLVLQGLALLP-----AATQGNKVLGKGGDVELTCTASQKSLQFHKNSNQKI 61
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 2 KILWFFLLVNAFRRVLSQVQLQWAGILKRPSETLTLCGGYGGFSFGYWSWIRQPG 61
Qy 62 LGNQ--GSFLTGPKELNDRADRSRL--WDQGNFLIKNLIKIESDT--YICEVEDQKE 116
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 62 KLEWIGELINSGITNNYPSLSKRVITISVDTSKQSLSLSSVNAADPAVYIC----- 114
Qy 117 EVQLLVFGITLANSSTHL---LOGQSLTLTLSPSSPS---VQCRSPRGKNIQGGKTL 169
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 115 --ARVITRASPCTDGRGMDVWGQGTITVTVSSGSAAPLFLPVLC----- 158
Qy 170 SVSQELDQSGTWTCTVLOKQKVEFKIDIVLAQ--KASSTVYKKEGQVFPPLAFT 228
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 159 -----ENSPSDTSSVAVGCLAQDFLPDSITFSMKYKNNSDISSTRG-----FPSVLR 205
Qy 229 VEKLTSGEELMMQAEASSSKSWITFDLNKKEVSVKRYTQDPKLGKKLPLHL--TLPQ 286
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 206 GGTAAATGQVLLPSKQV-----MGTDEHVCKVQHPNGKKEKNVPLPVIAELRP 255
Qy 287 ALPOYA-----GSGNLTALAEKTKLHGEVNLVNR-----ATOLQKNIACEVW 331
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 256 KVSVPFPRDGFPGNPKRSKLIQATGSPRQIQVSWLRBKGQVSGVTTDQVQAIAKES 315
Qy 332 GPTSPKMLSLKLENKEAKVSKREKPVVNLNPEAGMOCCLSDSGVILE--SNIKVLPW 390
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 316 GPTTKVSTLTITLKESD-----WL---SQSMFCRVDRHRLTFQONASSMCVDPQ 362
Qy 391 STPVPKSCDKHTKPCPCAPELLGSPVFLPRPKKOTLMISRTPEVTCVVVDVSHEDP 450
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 363 DTAIR-----VFALPPS--FASIFLTSTKTLCLVTLDTLTYD-- 397
Qy 451 EVKFMVYDQVEVHNAKTKPREEQYNSTYRVVSVLTVLIHQDMLNGEKYCKVSNKALPAP 510
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 398 STLTSTWNGEAVKTHNTNISHPATSAVGEASICEEDMNSGERFTCTVTHDLPSP 457
Qy 511 IEKTSKAKGQF--REPQVYTLPPSRDELTKNOVSLTCLVKGFPSPDIADVEMESNGQF-- 566
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 458 LKQITSRPKGVALHNPVDVLLPAPAEQLMLRESAITTCLVGFSPADVFMQMGQQLPS 517
Qy 567 ENNYKTPPVLD--SDGSFPLYSKLTVDKSRMOQGNVSCVMHEALNHTYQKSLSLSP 624
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 518 PEKVTSAPEPQAPGRYFAHSILTVEEEMNTGETYTCVVAHEALPNRYERTVDKST 577
Qy 625 G 625
      |
Db 578 G 578

RESULT 49
Q8WUX4 PRELIMINARY: PRT: 588 AA.
ID Q8WUX4:
AC 08WUX4:
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph.
RA Strausberg R.;
RL Submitted (Dec-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC019235; AAH19235.1; -.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.

```



```

RT receptor gene).";
RL Vopr. Vltu601. 40:100-102(1995).
DR EMBL, X87579; CAA60883.1; -.
DR EMBL, S79267; AAB35273.1; -.
DR PIR, I60082; I60082.
DR HSP, P01730; ICDY.
DR GO; GO:0004872; F:receptor activity; IEA.
DR InterPro; IPR007110; IG-like.
DR PROSITE; PS50835; IG_LIKE; 1.
KW Receptor.
FT NON_TER 71 71
SQ SEQUENCE 71 AA; 7844 MW; ASC9D84816135C86 CRC64;

Query Match
Best Local Similarity 98.6%; Score 357; DB 4; Length 71;
Matches 70; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NNRGVFRRLLLVLTQALLPAATQGNKVLGKKGDVETLTCTASQKKSIOFHMKNNSQIK 60
DB 1 NNRGVFRRLLLVLTQALLPAATQGNKVLGKKGDVETLTCTASQKKSIOFHMKNNSQIK 60
QY 61 ILGNQGSFLTK 71
DB 61 ILGNQGSFLTK 71

RESULT 52
Q7TOR1 PRELIMINARY; PRT; 587 AA.
AC Q7TOR1;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_Taxid=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=22341132; PubMed=12454917;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.,
RA "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RA Initiative.";
RT Dev. Dyn. 225:384-391(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.W., Hong L.,
RA Stalcup M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustun T.B., Toshiyuki S., Carninci P., Prange C.,
RA Bask S.S., Loughellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Rohak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Gilmwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalins D.E., Scherch A., Schein J.E.,
RA Jones S.J., Matra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]

```

```

RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RA Klein S., Strausberg R.,
RA Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL, BC056078; AAH56078.1; -.
KW Hypothetical protein.
SQ SEQUENCE 587 AA; 64933 MW; 646B5FD0FA414EBD CRC64;

Query Match
Best Local Similarity 9.8%; Score 333.5; DB 13; Length 587;
Matches 98; Conservative 51; Mismatches 118; Indels 65; Gaps 13;

QY 342 LKENKAKVSKREKPV-----WYLNPEAGMMQCL--SDSGOVL 380
DB 280 LKNGQTTEGVRVEPEVDKKRGVATSYLITRKEMDLDT---LYSCVGHASGS-LQ 335
QY 381 ESNIVLPTWSTPVEPKSCDTHTCPCPAPELLGGSEVFLPPPKQDTLMISRTPEVTC 440
DB 336 EKNMSKSLMCDTPIPTSI-----QVITIPPS-LESIFEKKSATLTC 376
QY 441 VVVDVSHDPEVKFNWY-VDGVFNAAKTKRREQYNS--TYRVSVLTVLHODMLNKE 497
DB 377 LVSNMANSDELRSISWFKSGTQETPLKTELGDATYNDNRITYSVKTTTVCADENNDK- 435
QY 498 YCKKSNKALPAPIEKTSKAKGQPREPQVYTLPPSRDELTKNQ-VSLTCLVKGFYPSDI 556
DB 436 FVCKVEHTELASMKVEFLFKKGEVNTPSYVFPPELSELSKREKATILCLVKGPSPEI 495
QY 557 AVEW--ESNGQPENNYKTP-----PYLSDGSFFLYSKLTVDSKRNQGNVFSQVMH 608
DB 496 FVKMLHKNEAVPKQYINTSINDELLPKGQSGKFFLYSLHTIDIKWDADSDSFSCVVG 555
QY 609 EALNHHYQKSLSSPG-----LQDDFC 632
DB 556 ESLPLQLQRSIDKSGKPTNVNVLVLDTC 587

RESULT 53
Q90544 PRELIMINARY; PRT; 684 AA.
AC Q90544;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Novel antigen receptor precursor.
OS Ginglymostoma cirratum (Nurse shark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Galeoidea; Orectolobiformes;
OC Ginglymostomatidae; Ginglymostoma.
OX NCBI_Taxid=7801;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=95183140; PubMed=7877689;
RA Greenberg A.S., Avila D., Hughes M., Hughes A., McKinney E.C.,
RA Flajnik M.F.;
RT "A new antigen receptor gene family that undergoes rearrangement and
RT extensive somatic diversification in sharks.";
RL Nature 374:168-173(1995).
DR EMBL, U18701; AAB48195.1; -.
DR PIR, S60266; S60266.
DR HSP, P01842; 2MCG.
DR GO; GO:0004872; F:receptor activity; IEA.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG_C1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; IG_6; 6.
DR SMART; SM00407; IG_C1; 4.
DR PROSITE; PS50835; IG_LIKE; 6.
DR PROSITE; PS00290; IG_MHC; 3.
KW Receptor; Signal.
FT SIGNAL 1 18
FT CHAIN 19 684
FT NOVEL ANTIGEN RECEPTOR.

```

SQ SEQUENCE 684 AA; 75224 MW; 2FF9D2071CDADFD CRC64;
 Query Match 9.7%; Score 330.5; DB 13; Length 684;
 Best Local Similarity 23.5%; Pred. No. 1.8e-16;
 Matches 158; Conservative 83; Mismatches 237; Indels 195; Gaps 29;
 QY 16 LALLPAATQGNKVNVLGKGDVTELTCTAS--QKSIQIOWH-KNSNQIKILNGSGFLTKG 72
 DB 145 VSLHSAATEEQRRA-----NRFQVLCLISGYYPENIAVSMOKNTYTI-----TSGFATTS 194
 QY 73 PSKLNDRADSRSL-----WDOGNFPLIKNLKIEDSDTYICEVEDQKEVQLLVFGL 125
 DB 195 PFKTSSNDFSCASLAKVPLQEWSSRS-----YISQVY----- 226
 QY 126 TANSDTLLQGSLLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCT 185
 DB 227 -SHSATSSNQKKEINST-----SEIAVLNRDP-----TYEIRIMDKSATLVCE 268
 QY 186 VLQONKKEFKIDIVVLAFOKASSIVYKKEGEQV--FSFPLATTEKLTGSGELMQA 242
 DB 269 VLSTVSAG-----VVSMMVNGKV--ENEGVQMEPTKMSGNOYLITRLTSSVEEW--- 317
 QY 243 ERASSSKSMITPDLKNKEVSVKRVYQDPKLOMGKKLPLHLTLPOALPOYAGSGNLTAL 302
 DB 318 -----QSGVEYTCSAKQDQSGSTPVVAKTRKARVEPTKP-HLRLLPSPBEIQSTG--- 366
 QY 303 AKTGKLHOEVLVWNRATOLQKLTCEVWGPTSPKLMLSKLENKEAKVSKREKPVWLN 362
 DB 367 -----SATLTCLIRGFYPRDKVSVS----- 385
 QY 363 PEAGMWQCLSDSGVLLSENIKVLPTM-----STVEPKSCDKHTHC--- 405
 DB 386 -----WQ-----KDDVSANVTNFTALBODLPTSTRSLNLTVEWKSQAK-YTCVAS 434
 QY 406 -PP-----CPABELGSPSVFLPFPKPKDPTLMISTEPTVTCVVDVSHED 449
 DB 435 HPPSGSTVKRYIRNOKVDCRQGTI---SVSLIKP-PREEIWTQGTATITVEIV---YSD 486
 QY 450 PE-VKFNMYVDGVEVNAHAKTREREOYNSTVRVSVLTVLHODMLNGEKYCKVSNKALP 508
 DB 487 LENIKVFQVNGVEKKGKGVETQNPMSGSKSTIVSKLKVMASEMSGTEVEYCLVEDSELP 546
 QY 509 APIETKISKAK-GQPREPOVYTLPPSRBEL-TKQVNSLTCLVKGYPBDIAYEWSNQ- 565
 DB 547 TPVKASIKKAVSQMHPKRYLLHPSTBEIDENATIMCATINHPAIIYVGMWANDTL 606
 QY 566 PENNYKTPPYLDSPFFLYSKLVVDSRMQGVFSCVWHEALH-----NHYTQK 618
 DB 607 LDGSRITQVDBEKSGSSGFVDRRLTLTAEMNSDTTYSCLVGHPSLNRDLIRSTNKSNGK 666
 QY 619 SLSLSPGLQDDET 631
 DB 667 PSSVNVSVVLSDT 679
 RESULT 54
 Q29027 PRELIMINARY; PRT; 99 AA.
 AC Q29027;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DB CD4, allele 1 (Fragment).
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OK NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN#1183; TISSUE=Blood;
 RX MEDLINE=93329116; PubMed=8335933;
 RA Gustafsson K., Germana S., Sundt T.M., Sachs D.H., LeGuern C.;
 RT "Extensive allelic polymorphism in an exposed region of the

RT miniature."
 RL J. Immunol. 151:1365-1370(1993).
 DR EMBL; X65629; CAA46583.1; -.
 DR PIR; I47131; S21461.
 DR HSSP; P01730; 1CDY.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003596; Ig_v.
 DR Pfam; PF00047; Ig; 1.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS50835; IG_LIKE; 1.
 FT NON_TER 1
 FT NON_TER 99
 SQ SEQUENCE 99 AA; 11170 MW; 40BF080699CF5DOC CRC64;
 Query Match 8.9%; Score 305.5; DB 6; Length 99;
 Best Local Similarity 60.2%; Pred. No. 9.3e-16;
 Matches 59; Conservative 16; Mismatches 22; Indels 1; Gaps 1;
 QY 32 KGGDTVELTCTASQKSIQFHWNKSNQIKILNGSGFL-TKGPSKLNDRADSRSLMDQG 90
 DB 1 KAGDLALPCHSSQKKNLPPNWKNSNOTKILGCHGSFWHTASVTELTSLDSKKNMWDHG 60
 QY 91 NFPLIKNLKIEDSDTYICEVEDQKEVQLLVFGLTAN 128
 DB 61 SFPLIKNLLEVTDGTYICEVEDKRIEQLLVFRLTAS 98
 RESULT 55
 Q72379 PRELIMINARY; PRT; 478 AA.
 AC Q72379;
 DT 01-OCT-2003 (Tremblrel. 25, Created)
 DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Hypothetical protein DKFZ686K04218 (Fragment).
 GN DKFZ686K04218.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OK NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Human rectum tumor;
 RA Bloecher H., Boecher M., Mewes H.W., Weil B., Amid C., Osanger A.,
 RA Fobo G., Han M., Wiemann S.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BX38066; CAD97996.1; -.
 KM Hypothetical protein.
 FT NON_TER 1
 FT NON_TER 478
 SQ SEQUENCE 478 AA; 51620 MW; 4AFCB541F3217CA1 CRC64;
 Query Match 8.9%; Score 302.5; DB 4; Length 478;
 Best Local Similarity 24.0%; Pred. No. 1.5e-14;
 Matches 132; Conservative 66; Mismatches 174; Indels 179; Gaps 27;
 QY 171 VSQLELDQSG-----TWCTVLOKQKVEFKIDIVVLAFOKASSIVYKKEGEQV 220
 DB 17 LSQVQLQESGPRGLYKPSQTLSTCTV-----SGGSI----- 47
 QY 221 FSPFLATTEKLTGSGELMQAERASSKSMITPDLKNKEVSVKRVYQDPKLOMGKKLPL 280
 DB 48 -----GSGDYFW-----SMI-----RQAP-----GRGLEW 67
 QY 281 HLTLPOALPOYAGS-----GNLTALFAKQKLGHEVY-----LVWNRATOLQKN 325
 DB 68 -----MGYIYSGSTYVNPESLRSLSTIDTSKNQSLRLNSLTADTAVYFCARGVGLG 122
 QY 326 LTCEVWG-----PTSPKLM-LSLKLENKEAKVSKR-----EKPVVVLNPEA 365
 DB 123 TAPDIWQGVVVVYSSASPTSPKVPFLSLDSTQDGNVVVACLVQGFPPQEPPLSVTMS 182
 QY 366 GM-----WQCLSDSGVLLSENIKVLPTWSTPVEPKS--CD-KHTTCTP-----PCPA 410

Db 183 GQNTARNPPSODASGDLTYTSSQLTLPATQCP-DGKSVTCHVKHYTNPQDVTVPCHV 241
QY 411 PELGGSPVFLPPKPF-----KDTLMISTPVTCTVVDVSHEDPEVKRWV 458
Db 242 P-----PPPCCHPRLSLHRPALEDLLGSEANLTCLTGL-RDASGATFTWP 289
QY 459 DGEVNAKTKPREEOYNSTYRVSVLTVLHODMLNGKEYCKVSNKALPAPIEKTISKA 518
Db 290 SSGK--SAVQGPFRRLCCGYSVSVLPGCAQPMNGETFTCTAAPLKTPLTANITKS 347
QY 519 KQGPPEQVYTLPPSDELTKQV-VSLTCLVKGFPYSDIAVWESNGQ--PENNYKTPP 575
Db 348 -GNTFRPEVHLPPPESEELALNELVTLTCLARGSPKDVLRVLQSGQLPREKYLTMAS 406
QY 576 VLD-SGQ--SFLYSLKLTVDKSRWQGNVPSGVMHEALHNHTOKSLSPG----- 625
Db 407 RQPSGCTTFAVTSLIRVAEDMKGDFTSCVSHALPLAFTOKTIDRLAGKPTHNV 466
QY 626 ----LQDDETC 632
Db 467 SVMMAEVDGTC 477

RESULT 56

Q96K68 PRELIMINARY; PRT; 494 AA.
ID 096K68
AC 096K68
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
DE Hypothetical protein FLJ14473.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Isegaki T., Ota T., Hayaishi K., Sugiyama T., Otsuki T., Suzuki Y.,
Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
Wagatsuma M., Hosokiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
Takahashi M., Chida Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
Ninomiya K., Iwayanagi T.;
RT "NEO human cDNA sequencing project";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027379; BAB55072.1; -
DR PIR; S21205; S21205.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG_4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
KW Hypothetical protein.
SQ SEQUENCE 494 AA; 53088 MW; 9A1D7AB5AE64C0E CRC64;

Query Match 8.7%; Score 298.5; DB 4; Length 494;
Best Local Similarity 21.9%; Pred. No. 3.1e-14;
Matches 140; Conservative 85; Mismatches 204; Indels 209; Gaps 28;

QY 30 LGKGGDTVELTCTAS--QKKSIOFWH-----KNSNQIKLNGQSF-----TKGPSKL 76
Db 30 LVKPGGSLRLSCAASLFSFTYAMNVRQAPGKLEWVSISRSRDIYIRDVSKGRFTI 89
QY 77 NRPADSRRLMDQGNPFLIIKNLKIEDSTYICEVDQKEVQVLVFGITANSPTHLOG 136
Db 90 S-RDNKNKSLYLVQW-----SLKVDLTAVYICARSCNGAI---CYGFSF-----WG 132
QY 137 QSLTTLSPGSSPSPVQCRSPKGNIGGKTLVSQLELQDSGTWTCTVLQNKQKVEPK 196

Db 133 QGTLVTVASASPTSP-----KVFLPILCSTQPDG-----NVV 164
QY 197 IDIVLAF--QKASIVYKKEGEQEF--SFLAFLVEKLTGSGELMQAERASSSKSMT 253
Db 165 IACLVQGFPPQPLPSVTSESQGVYARNF-----PSQDASGDLV-----TSSQ---- 210
QY 254 FDLKKKEVSVKRVTDQPKLQMGKULPLHITLTPQALPQYAGSNGLTALAEAKTKLHQEVN 313
Db 211 -----LTLF----- 214
QY 314 LVWKRATQ--LQKNTLCEWNGPTSPKMLSLKLENKAKVSKREKRVWLNBEGMOC 371
Db 215 ----ATQCLAGKSVTCHVKHYTNPQDVTVP----- 241
QY 372 LSDSQVLLNESIKVLPWSTPVEKSCDKHTTCPCAPELLGSPVFLPPKRDITM 431
Db 242 -----CPVSPSTPPT-----SPSTPPTSPSC-HPRLSLHPALD-L 279
QY 432 ISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVNAKTKPREEQYNSTYRVSVLTVLHOD 491
Db 280 LQSEANLTCLTGL-RDASGVTFTWTPSSGK--SAVQGPFRRLCCGYSVSVLPGCAEP 336
QY 492 WLNGKEYCKVSNKALPAPIEKTISKAGQPREPOVYTLPPSDELTKQV-VSLTCLVKG 550
Db 337 WNHGKTFCTTAAYPESKTPPLTATLSKS-GNTRPREVHLPPPESEELALNELVTLTCLARG 395
QY 551 FYPSDIAVWESNGQ--PENNYKTPPVLV-SDG--SFLYSLKLTVDKSRWQGNVPSGS 605
Db 396 FSPKVLVLRWLQSGSELPREKLTWASROBPSQGTTFATVSLIRVAEDMKGDFTSCM 455
QY 606 VNHREALHNHYTOKSLSPG-----LQDDETC 632
Db 456 VGHREALPLAFTOKTIDRLAGKPTHNVSVMAEVDGTC 493

RESULT 57

Q96KX8 PRELIMINARY; PRT; 496 AA.
ID 096KX8
AC 096KX8
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Stiusberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC016369; AAH16369.1; -
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG_4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
KW Hypothetical protein.
SQ SEQUENCE 496 AA; 53391 MW; D346929849040D69 CRC64;

Query Match 8.7%; Score 296; DB 4; Length 496;
Best Local Similarity 23.5%; Pred. No. 4.9e-14;
Matches 133; Conservative 72; Mismatches 205; Indels 156; Gaps 22;

QY 124 GITANSPTHLOGGSLTTLSPGSSPSPVQ---CRSRGNIGGKTLVSQ-----E 175
Db 29 GLVKSST-----LSLTCTVSGSGSISSSYWGWIRQPGKLEIANITYSGITYNPS 83
QY 176 LQDSGTWCTVVLQNKQKVEFKIDIVLAFQKASIVYKKEGEQVFSPLAFTVEKLTGS 235

```

Db      84 LKSRRTISVDISKNQ-----LSLKRVSTADTAVY-----FCANHGVSRSRTGA 123
Qy      236 GELMQOABRASSKSKWTFDLKNKEVSVKRVTPDPLQMGKKLPLHLTLPLQALPOLYAGSG 295
Db      130 IDYMQOGLTVTVSSASPT-----SPKV-----FPLSLCSTCP-----DG 163
Qy      296 NLTLTL-----EAKTGKLGQEVNLVWRATQ--LQ 323
Db      164 NVVILCVLQGFPPQEPRLSVTMSSESGGVTAARNPPSQDASGDLYTSSQLTLPLATQCLAG 223
Qy      324 KNLTECVMGPTSPKMLSLKLENKEAKVSKREKPVWVLNPEAGMMQCLLSDSGVLLBSN 383
Db      224 KSVTCHVKGHTNTPSQDVTVP----- 243
Qy      384 IKVLPFTWSTPVEPKSCDKTHTCPCPAPRLGSPVFLFPPKQDTLMSRPEVTCVVV 443
Db      244 ---CVPSPPTPP-----SPSTPTTPSPSCC--HPRLSLHRPALDD--LLGSEANLTCTLT 293
Qy      444 DYSHEDPEKFMWYVDGVEVHNAKTRPREQNSTGRVSVLTVLHQDLNKEVKKYS 503
Db      294 GILRDASGTFWTPSSSGK--SAVQGPPEPRLCGCYSVSVLPGCAEPNNHGKTFCTTAA 350
Qy      504 NKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQ--VSTLCLVKGPSDIADVEMES 562
Db      351 YPESTKPLPLATLSKS-GNFRFEVHLPPRSEELANLVLTLCTLARGSPDPVDLVRWQ 409
Qy      563 NGQ--PENNYKTPPYLD-SDG--SFFLYSKLTVDKSRMQGQNVSPCAVMEALHNHYTQ 617
Db      410 GSGELPREKYLFWASRGQPSGTTTFVAVTSILRVAAEDWKKGDPTFCMGVHALPLAFIQ 469
Qy      618 KSLSLSPG-----LQDLETC 632
Db      470 KTIDRLAGKPTHVNVSVVMAEVDGTC 495

RESULT 58
OBN5K4
ID      OBN5K4      PRELIMINARY;      PRT;      499 AA.
AC      OBN5K4;
DT      01-OCT-2002 (TREMBLrel. 22, Created)
DT      01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT      01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE      Hypothetical protein.
OS      Homo sapiens (Human).
OC      Eukaryota; Euteleostomi; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Euteria; Primates; Catarrhini; Homidae; Homo.
RX      NCBI_Taxid=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Blood;
RA      Struhsberg R.;
RL      Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL; BC032249; AAH32249.1; -
DR      InterPro; IPR003599; Ig_
DR      InterPro; IPR007110; Ig-like
DR      InterPro; IPR003597; Ig_C1
DR      InterPro; IPR003006; Ig_MHC
DR      InterPro; IPR003596; Ig_V
DR      Pfam; PF00047; Ig_4
DR      SMART; SM00409; IG; 4
DR      SMART; SM00407; IG1; 2
DR      SMART; SM00406; IG; 1
DR      PROSITE; PS50835; IGV; 1
DR      PROSITE; PS00290; IG_MHC; 1
KW      Hypothetical protein
SQ      SEQUENCE 499 AA; 53376 MW; 93A5C89582054F32 CR664;

Query Match      8.6%; Score 295; DB 4; Length 499;
Beet Local Similarity 25.8%; Pred. No. 5,9e-14;
Matches 133; Conservative 61; Mismatches 202; Indels 120; Gaps 25

219 VEFSPPLAFTYEKLTG-----SGELMQOABRASSKSKWTFDLKNKEV 261
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db      1 MEFGLSWFLVALILNGQCEVQLVBSGGGVARPGSL-----RLSCATSGFTPD-----DS 51
Qy      262 SVKRVITQDPKLOMGKLP--LHLTLPOLPYAGS--GNLTALAEKTKLHOEVLVVM 317
Db      52 GASWROAP-----GKLEWVSSINMNGSGSTNVADSVKGFRTISRDNAKNSLYLQNNSLRV 1070
Qy      318 RATQIQ-----KNLTC-----EWVG-----PTSPKLM--LSLLENKKEAK 350
Db      108 EDTALYYCARDPPTYKCSGGSCGCGYIMDVWKGKTTVTYSSASFTSPVFFLSLSTQPDGN 167
Qy      351 VSKR-----EKRVWVILNPEAGM-----WQCLLSDSGOVLLBSNIKLP-TWSTPVE 359
Db      168 VVIACLVGFFPQFQELSLVSWSSGGGVTAARNRPPSQDASGDLTYTSSQTLTLPATQCLAGK 227
Qy      366 PKSCD-KYHTCP-----PCPAPELIGGPSVLFPFKPK-----DTLMTS 433
Db      228 SVTGVCHKYHTNPSQDVTPCVPSPPTPTPTSPST-PTPTSPSCCHPRLSLHPALBDLLG 266
Qy      434 RPEVTCVAVDVSHDEPEKFMWYVDGVVNAKTPREBOYNSTRVYSLTVLHQDWL 493
Db      287 SEANLTCTLTG--RDASGVTFMTPTSSGK--SAVQGPDRDLGGCYSSVSLPGCAEPWN 343
Qy      494 NGEKYCKVSNALPAPIEKITSKAGQPREQVYTLPPSBDLTAKNO-VSLTCLYKGFY 552
Db      344 HKCTTCCTAAVBSKTPPLATISTKS-GNFRFREVHLPPPSBELANELVTLTCLARGFS 402
Qy      553 PSDIAVESNGQ--PENNYKTPPYLD-SDG--SFFLYSKLTVDKSRMOQGVNPSCSVM 607
Db      403 PKDVIVLRMLQSGQLPREKXYLTWASRQSPSQGTTTFAVYSILRAVEDWVKGDTFSCMVG 462
Qy      608 HEALHNHYTKSLSLSPG-----LQDENC 632
Db      463 HEALPLAFTQKTRIDRLAGKPTHVNVSVMAEVDGTC 498

RESULT 59
Q8NCL6
ID      Q8NCL6      PRELIMINARY;      PRT;      493 AA.
AC      Q8NCL6;
DT      01-OCT-2002 (TREMBLrel. 22, Created)
DT      01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT      01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE      Hypothetical protein FLJ90170.
OS      Homo sapiens (Human).
OC      Eukaryota; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eulalia; Primates; Catarrhini; Hominiidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Mammary gland;
RA      Itoigai T., Ota T., Nishikawa T., Hayaeshi K., Otsuki T., Sugiyama T.,
RA      Suzuki Y., Nagai K., Sugano S., Iehli S., Kawai-Hio Y., Saito K.,
RA      Yamamoto Y., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
RA      Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotaka S., Sasaki N.,
RA      Hattori A., Okumura K., Iwayanagi T., Niimiya K.;
RT      "NEO human cDNA sequencing project.";
RL      Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR      InterPro: IPR003599; Ig_
DR      InterPro: IPR007110; Ig-like.
DR      InterPro: IPR003597; Ig_cl.
DR      InterPro: IPR003006; Ig_MHC.
DR      InterPro: IPR003596; Ig_V.
DR      Pfam: PF00047; Ig_4.
DR      SMART: SM00409; Ig_4.
DR      SMART: SM00407; IG1.2.
DR      SMART: SM00406; IG_1.
DR      PROSITE: PS50835; IG_LIKE; 4.
DR      PROSITE: PS00290; IG_MHC; 1.
KW      Hypothetical protein.
SQ      SEQUENCE 493 AA, 53224 MW, 126CD7E094777101 CRC64;

```



```

Best Local Similarity 23.0%; Pred. No. 9.8e-14;
Matches 148; Conservative 80; Mismatches 203; Indels 212; Gaps 31;

QY 25 GKKVVLGKGDPTVELCTAS--OKSIOFHMKXNSQ-----IKIINGOG-----SFLRK 71
DB 27 GGGVVL--PGSIRLSCAASGRFRFYDMHWARQSGEGLWVALIWTGTTTYSDSVK 84
QY 72 GSKLNDRAISRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEVQLLVGLTANSDPT 131
DB 85 GLRTVS-RDNYKNTLY-----LEMKSLGAEDTAIVYC-ARDQ---GVAIGVF-----D 128
QY 132 HLLQGSLLTLSPSPSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTYLQONK 191
DB 129 HMGQGLVTVSSASP--TSP-----KVPFLSLCSTQTPDG----- 160
QY 192 KVEPKIDIVLAF--OKASIVYKKEGEVGF-SFPLATVEKLTGSGELMMQAEASRS 248
DB 161 --NVVIACLVQGFPOEPLSVTWSGGQVTAIRNF-----PSQASGLY-----TSS 208
QY 249 KSMITFDLKNKEVSVKVTQDPKLQMGKLLPLHLTPQALPOYAGSGNLTLLAEAKTKL 308
DB 209 Q-----LTLF----- 213
QY 309 HOEVLNLVNRATQ--LQKNLTCEVWGPTSPKLMLSIKLENKAKVSKREKPVWVLPENXG 366
DB 214 -----ATQCLAGKSVTCHVKHYTNPSQDVTVP----- 240
QY 367 MMQCLSDSGOVLBSNINVLPTWSTPVEPKSCDKHTPCPCAPBELLGSPVFLPPKP 426
DB 241 -----CPVSTPPPP-----SPSTPPSPSCC-HPRLSLHRPL 274
QY 427 KDTLMISRTPEVTCVVDVSHEDPEVKFMYVGVGVHNAKTRPREQYNSYRVVSVLT 486
DB 275 ED-LILGSEANLTCTLTGL-RDASGVTFTWTPSSGK--SAVQGPREDLCGCYSVSVLP 330
QY 487 VLDHQLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQ-VSIT 545
DB 331 GCAEPNHNKFTCTCTAAYPESKTPLATLTSKS-GNTPREPVHLPPSEBALNLELVLT 389
QY 546 CLVKGFPYPSDIAVEMESNQ--PENNYKTTTPVLD-SDG--SFPLYSKLTVDKSMQGN 600
DB 390 CLARGSPDPVLVVRWLQGSQELPREKYLTMASRQEPSSQGTTFFAVYSILRVAEDMKGD 449
QY 601 VFSCSVMEHALNHNHYTKSLSPG-----LQDETC 632
DB 450 TTSQVNGHEALPLAFQKTIIDRLAGKPTVNVSVMAEVDGTC 492

RESULT 60
Q9NRP6 PRELIMINARY; PRT; 416 AA.
AC Q9NRP6;
DB 01-OCT-2000 (TREMBLrel. 15, Created)
DB 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DB 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DB Immunoglobulin heavy chain variant (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Pluvine R., Estivill X., Becarceller M., Sumoy L.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Aufreay C., Amosge W., Ballabio A., Estivill X., Gibson K.,
RA Lehrach H., Poustka A., Lundberg J.;
RT "The European IMAGE consortium for integrated Molecular analysis of
human gene transcripts."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL389978; CAB97534.1; -.
DR HSP; P01789; IMCP.

```

```

DR InterPro; IPR007110; IG-1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG_4.
DR SMART; SM00405; IGv_1.
DR PROSITE; PS00835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 2.
FT NON TER
SQ SEQUENCE 416 AA; 44786 MW; 8C41708BB8A84687 CRC64;

Query Match 8.5%; Score 291; DB 4; Length 416;
Best Local Similarity 28.0%; Pred. No. 9.2e-14;
Matches 103; Conservative 48; Mismatches 137; Indels 80; Gaps 18;

QY 329 EYWG-----PTSPKLM-LSLKLENKAKVSKR-----EKPVWVLPENAGM- 367
DB 64 DVMGGTTVTVSSASPTSPKFPPLSLDSTPDQNVVACLVQGFPOEPLSVTWSGON 123
QY 368 -----WQCLSDSGOVLBSNINVLPTWSTPVEPKS--CD-KTHTCP-----PCPABEL 413
DB 124 VTARNFPPSQASGDLVTTSSQLTLPATQCP--DGKSVTCHVKHYTNPSQDVTVP-- 180
QY 414 LGPSVFLPPKP-----KDTLMISRTPEVTCVVDVSHEDPEVKFMYVVDGV 461
DB 181 -----PPPCCHPRLSLHRPALBDLLGSEANLTCTLTGL-RDASGATFTWTPSSG 230
QY 462 EYHNAKTRPREQYNSYRVVSVLTVHQMILNGKEYCKVSNKALPAPIEKTISKAKG 521
DB 231 K-SAVQGPREDLCGCYSVSVLPFGCAQPMNHGTFCTAHPKLTPLTNITKS-GN 287
QY 522 PREPOVYTLPPSRDELTKNQ-VSLTCLVKGFPYPSDIAVEMESNQ--PENNYKTTTPVLD 578
DB 288 TPREPVHLPPSEBALNLELVTLCLARGSPKDVLRWLQGSQELPREKLTWASRQE 347
QY 579 -SDG--SFPLYSKLTVDKSMQGNVFCSCVMEHALNHNHYTKSLSPG----- 625
DB 348 PSQGTTFFAVYSILRVAEDMKGDTFSCVNGHEALPLAFQKTIIDRLAGKPTVNVSV 407
QY 626 -LQDETC 632
DB 408 MAEVDGTC 415

RESULT 61
Q9BRV0 PRELIMINARY; PRT; 500 AA.
AC Q9BRV0;
DB 01-JUN-2001 (TREMBLrel. 17, Created)
DB 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DB 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DB Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC005951; AAH05951.1; -.
DR HSP; P01789; IMCP.
DR InterPro; IPR007110; IG-1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG_4.
DR SMART; SM00406; IGv_1.
DR PROSITE; PS00835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
KW Hypothetical protein.
SQ SEQUENCE 500 AA; 54154 MW; 0A9BF43F2A3CC6D9 CRC64;

Query Match 8.4%; Score 288; DB 4; Length 500;

```

Match	Local similarity	23.1%	Pred. No. 2e-13	Matches 204	Indels 188	Gaps 27
138	Conservative	68	Mismatches	204	Indels 188	Gaps 27
Qy	118	VQLVFGILT-ANSDTHLQGGSLTLTLESPPGSSPSSVOCRSRGRKNIQGGKTLVSQLEL	176			
Db	7	ILFLVAAATGAGOSQVHLVOSGAEVMS-----PGASVRSVCKT-----	43			
Qy	177	QDSGWTCVTVLQNKQKVFKIDIVLAPQKASSIYYKKKGQEVRSFPLAFVETLTSQG	236			
Db	44	--SG-----YAHHTYSII-----WRCAPGCG	63			
Qy	237	ELMWOAERASSKSWITPDLKNKEVSVKRVKTODPKLQMGKLLPLHLPLQALPOYAGSGN	296			
Db	64	LEW-----MGWISSPSDN-----TRFAKRP-----QGR	86			
Qy	297	LTLALFATGTLKHQV-----NLVWRARQLOGLNLTC-----EYWG-----	332			
Db	87	VTLTDTISTSTVYMLRSLRSDDTAVVYCARRYCSYSSCQNDYYYYYMDWGGKTTVTVS	146			
Qy	333	---PTSPKLM-LSLKLKNEKKAQVSR-----EKPPVNLNPEAGM-----WQCLSD	374			
Db	147	SASPTSPKRVPLSLCSTQPDGNNVLAICLVQGFPPPEPLSVTMSBGGQVTARNFPSQDA	206			
Qy	375	SGQVLLBSNIKVL-P-TWSTPVEPKSCD-KTHTCP-----PCPABELLGGBSVFLPPKP	426			
Db	207	SGDLVYTSQSULTPLPATQCLAGSKVYCHVMHYTNPQSDVTPCPVPSTPTPTPSPT-PTPT	265			
Qy	427	K-----DTLMTSRPEVTCVVVDVHEDREPEYKFNMYVVDGVENHNKTKPR	471			
Db	266	SPSCCHPLSLRHPALIEDLLGSEANLITCTLTGL-RDASGVYFTFTPSSGK--SAVQGRP	322			
Qy	472	EEQYNSTYRVVSVLTVTHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLR	531			
Db	323	DRDLGCGSVSVSLVSGCAEPMNHGKTFCTIAAYPESKTPRLPATLSKS-GNTPRPREVHLR	381			
Qy	532	PSRDELTKNQ-VSLTCLVKGFIPPSIDIAEWESNGQ--PENNYKTTTPVLVD-SDG--SFPL	585			
Db	382	PPSEELALNELVTLTCLARGSPSPKQVLRMLQGGSELPREKYLTVASSROEPPQGTTPPAV	441			
Qy	586	YSKLTVDVSRMQQGNVFCFSVWHEMLAHNHYTKSLSPG-----LQLDETC	632			
Db	442	TSILRVAEMDKKQGTPTFCSCMGHEALPLAFTQETIDRLAGKPTHNVSVMAEVDGTC	499			
RESULT 62						
Q96DK0	PRELIMINARY;	PRT;	496	AA.		
AC	Q96DK0;					
DT	01-DEC-2001 (TREMBLrel. 19, Created)					
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)					
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)					
DE	Hypothetical protein FLJ25298.					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.					
OX	NCBI_TaxID=9606;					
RN	(1)					
RP	SEQUENCE FROM N.A.					
RC	TISSUE=Gastric mucosa;					
RA	Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,					
RA	Hocuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,					
RA	Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,					
RA	Suzuki Y., Hata H., Nakagawa K., Mizuno S., Morigata M., Kawamura A.,					
RA	Sugiyama T., Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A.,					
RT	Kawakami B., Nagai K., Isogai T., Sugano S.;					
RL	"NBD0 human cDNA sequencing project."					
SR	Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.					
DR	EMBL; AK058027; BAB71633.1;					
DR	InterPro; IPR007110; IG-like.					
DR	InterPro; IPR003006; IG_MHC.					
DR	InterPro; IPR003596; IG_v.					
DR	Pfam; PF00047; IG; 4.					
DR	SMART; SM00406; IGV; 1.					

DR	PROSITE	PS50835; IG LIKE; 4.
DR	PROSITE	PS00290; IG_MHC; 1.
KM	Hypothetical protein.	
SQ	SEQUENCE	496 AA; 53532 MW; C72EE1E247C86FED CRC64;
Query Match	8.4%; Score 287.5; DB 4; Length 496;	
Best Local Similarity	23.2%; Pred. No. 2,26-13;	
Matches	136; Conservative 66; Mismatches 193; Indels 191; Gaps 27;	
OY	129 SDTHLLD-QGSGLTTLSPSSPSSVOCRSBRKNIOGKTLVSQLELDSCGTWCTVL	187
Dd	19 SQVHLVSGAGELKM-----PSSVKVSC-----	41
OY	188 QNOKKVEFKIDIVLAFQKASSIYYKKGEQVERSPFLAFT-VKLTGSGELMWQAERAS	246
Dd	42 -----KASAMFRS-----YATTWRAQGCGQLW-----	66
OY	247 SSKSMITFDLKNNKEVSVKRYTDPKLMGKPLRLHTLPALPOLYAGS--GNLTALAEAK	304
Dd	67 -----MGGIIPNF-----GAPVYAQNFODRVITISADDS	94
OY	305 TGTKHQEVNLVMMAQTOLQ---KNLT-----CEWG-----PTSPKLM-L	340
Dd	95 TTVTVMELTSLTFEDTAFFYCGRGTLTYGGSYYYLQHWGQGLTVLVTSASPTSPPKVPL	154
OY	341 SLKLNEEAASKR-----EKPVVLINPEAGM-----WCCLSDSCGVLLSENIKY	386
Dd	155 SLGCTPGDNVVILACLVQGFPPGPPLSVTWSESQGVATANFPSPGDASDLVYTSSQLT	214
OY	387 LP-TWSTPVEPKSD-KTHTCP-----PCPAPELLGSPVFLEPPKPK-----	427
Dd	215 LPAIQCLAGSKSVTHVGHGYTNPSODVTPCPVPSPPTPTPSPST-PPTPSPSCCHPRLSLH	273
OY	428 ----DTIMISTRPEVTCVVVDVSHEDDEVKFNWVVDDEVVHNAAKTRECYNSTYRWVS	483
Dd	274 RPALEDLLIGSEANULTTTLTGL-RDAAGVFPTWPSSGK--SAVGFPDDNDLCGCYSVS	330
OY	484 VLVTLVHDWLNGKEYKKCKVSNKALPADIEKTISSAKAQPREPOUYTLPPSRDELTXNQ-V	542
Dd	331 VLPSCAPEPMNHGKFTCTTAAYPESKTEPLTATLSKS-GNTFRPEVHLLPPSEBELIALNELV	389
OY	543 SLTGAVGFYPSDIADVEMESNGQ--PENNYKTTTPRVLD-SDG--SFPLYSKLTVDKSRMQ	597
Dd	390 TLTLTAGFSFKDVLVWMLGSGQLPPEKXYLTWASROEPQGTTFVAVTILIRVAEDMK	449
OY	598 QGANFSCSVMHAEAHNYTKSLSLSPG-----LQLDETCC 632	
Dd	450 KGRIFSCMVGHREALPLAFTQXTIDRLAGKPTHVNVSVMAEVDCTC 495	
RESULT	63	
ID	Q72374	
AC	Q72374;	
DT	01-OCT-2003 (TREMBLrel_25, Created)	
DT	01-OCT-2003 (TREMBLrel_25, Last Sequence Update)	
DT	01-OCT-2003 (TREMBLrel_25, Last Annotation Update)	
DE	Hypothetical protein DKFZP686C02218 (Fragment).	
GN	DKFZP686C02218.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxId=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Human rectum tumor;	
RA	Bloecher H., Boecher M., Mewes H.W., Weil B., Amid C., Oeanger A.,	
RA	Fobo G., Han M., Wiemann S.;	
RL	Submitted (JUN-2003) to the EMBL/GenBank/DDB databases.	
DR	EMBL; BX58077; CAD98001.1; -.	
KM	Hypothetical protein.	
FT	NON TER	
SQ	SEQUENCE 492 AA; 53776 MW; 1E7A57560F0CA74B CRC64;	

Query Match 8.4%; Score 286; DB 4; Length 492;
Best Local Similarity 27.9%; Pred. No. 2.8e-13;
Matches 102; Conservative 47; Mismatches 137; Indels 80; Gaps 18;

QY 331 WG-----PTSPKLM-LSLKLENKAVSKR-----EKPVWVLPBAGM--- 367
DB 142 WGGGLTVTVSSAPSPKVFPLSLDSTPDQGNVVAQLVQGFPPQEPPLSVTWSESGQVNT 201
QY 368 ---WQCLSDSGQVLLSNIKVLPTWSTPVEPKS--CD-KHTTCP-----PCPABELLG 415
DB 202 AANFPSPDASGDLTYTSSQLTLPATQCP-DGKSYTCHVKHKTNPDSQDVTVCPPV---- 256
QY 416 GRSVFLPPKP-----KDTLMTSRTPBVTGVVDVSHEDPEVKFWMYVDGVEV 463
DB 257 -----PPPCCHRLSLHRALEDLLIGSEANLTITGL-RDASGATFTWTSBGK- 307
QY 464 HNAKTRPREOYNSTYRVSVLTVLHQMVLNGEKYCKVSNKALPAPIEKTISKAKGQPR 523
DB 308 -SAGQPERDLGGCVSVSVLPGCAQPTHTGETFTCTAHPELKTPLTANITKS-GNTF 365
QY 524 EQQVYTLPPSRDELTKNQ-VSLTCLVKGFPYPSDIAVESNQ--PENNYKTTPVLD-S 579
DB 366 RBEVHLPPSEELALNELVLTCLAKGFPSPDVLVWLGSOELPREKYLTMASROEPS 425
QY 580 DG--SFFLYSKLTVDKSRWQGNVFSQVMEALNNHYTKSLSPG-----L 626
DB 426 QCTTFVAVTSLRVAEDMKGDTFSCWGHFALPLAFTOKTIDRLAGKPTHVNVSVMA 485
QY 627 QLEDT 632
DB 486 EVDGTC 491

RESULT 64
Q91WR1 PRELIMINARY; PRT; 488 AA.
ID Q91WR1
AC Q91WR1;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DT Hypothetical protein.
GN IGH-VJ558 OR A1893585.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Kidney;
RC Struoberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013539; AAI13539.1; -.
DR MGD; MGI:96486; Igh-VJ558.
DR InterPro; IPR007110; IG-1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003586; IG_V.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KW Hypothetical protein
SQ SEQUENCE 488 AA; 52964 MW; F12068460B400B9D CRC64;

Query Match 8.3%; Score 282; DB 11; Length 488;
Best Local Similarity 24.5%; Pred. No. 5.6e-13;
Matches 126; Conservative 78; Mismatches 198; Indels 112; Gaps 24;

QY 147 PSSSPSVQCRSPRGKNI-----QCGKTLG-VSGLDSDSGTWTTVQLQNKVEF 195
DB 33 PSASVLSCKA-SGYITIDYVVMWQSHGKSLWGLDINPNVNGT-----SYNOK---F 83
QY 196 KDIIVLARGKASIVYKKEGSEVERFPLAFTVEKLTSGLWQAERASSSKSMITPD 255

DB 84 K-GKATLVVDKSSSIAYWQ-----LNNTLSDSDSAVYCARGQVYVSFSYD 128
QY 256 LKN-KEYSVKRYVTDPLQMGKPLH-LTLPLALP-----QYAGSNTLAL 301
DB 129 RGDWGGQGLTVTVAEP---AREPTIPLTFPPQALSSDPVIGCLIHDPFPGTMYNTW 184
QY 302 EAKTGKHQEVNLVVMRATOLQKULTCFVWGPTSPKMLSLKLENKAVSKREKPVWL 361
DB 185 -GKSGKDTVYVFPALASGGRTYMSQLTPA-----VE 218
QY 362 NPEAGMOCLLSDSGQVLLSNIKVLPTWSTPVEPKSCDKHTTCPAPABELLGPVFL 421
DB 219 CPBGSSVVCVSHQSDNPQELVNVNCPGICSP-----TTPPPSCQ---PSLSL 264
QY 422 FPPKQDTLMTSRTPBVTGVVDVSHEDPE-VKRWYVDGVEVHNAKTRPREOYNST-- 478
DB 265 QRPALD-LLLGSDASITCTINGL-RDPEGAVFW-----EPSTGDAVQKKAQVNSCG 316
QY 479 -YRVSVLTVLHQMVLNGEKYCKVSNKALPAPIEKTISKAKGQPREOYVTLPPSRDEL 537
DB 317 CYSVSSVLPGBAERNNSASFTCTVTHPESDT-LGTIAKYVNTVFPQVHLLPPSEEL 375
QY 538 TKNQ-VSLTCLVKGFPYPSDIAVESNQ--PENNYKTTPVLDSDG--SFFLYSKLTV 591
DB 376 ALNELVSLTCLVRAFNPKVLYRWLHGNEELSPSYLVFEELKERGEGATTYLVTSVLRY 435
QY 592 DKSRWQGNVFSQVMEALNNHYTKSLSPG 625
DB 436 SAEIWKQGDQYSCWGHFALPMNFOTKIDRLSG 469

RESULT 65
Q9UP60 PRELIMINARY; PRT; 384 AA.
ID Q9UP60
AC Q9UP60;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE SMC73 protein.
GN SMC73.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zheng S., Cao J., Cao W., Cai X., Geng L.;
RT "Identification and characterization of SMC73, a gene which is down-regulated in colorectal cancer";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF067420; AAC19365.1; -.
DR HSP; P01825; 7FAB.
DR InterPro; IPR007110; IG-1like.
DR InterPro; IPR003597; IG_C1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00407; IGC1; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
SQ SEQUENCE 384 AA; 40947 MW; BA7ADC3CA5A9DD48 CRC64;

Query Match 8.2%; Score 281.5; DB 4; Length 384;
Best Local Similarity 26.8%; Pred. No. 4.3e-13;
Matches 93; Conservative 49; Mismatches 132; Indels 73; Gaps 13;

QY 305 TGKHHQEVNLVVMRATQ--LQKNLTCEVWGPTSPKMLSLKLENKAVSKREKPVWL 362
DB 91 SGLDLYTSSQLTLPATQCLAGSVTCHVKHKTNPDSQDVTVCPPV----- 131
QY 363 PEAGMOCLLSDSGQVLLSNIKVLPTWSTPVEPKSCDKHTTCPAPABELLGPVFL 422
DB 132 -----CPVSTPTPTP-----SPSTPTPTPSPSC-HPRSLH 161

```

Qy 423 PPKKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKRREQDYNSTYRVV
    |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 162 RPALED-LLLGSEANLTCTLTGL-RDASGVFTWTPSSGK--SAVQGPERRDLGGCVSVS 217

Qy 483 SVLVTVLHODWLNKGEYKCKVSNKALPAPIEKTISAKAQPREPOVYTLPPSRDELTKNQ- 541
    |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 218 SVLPQCAEPWNGKFTCTAAVPSKPTLTATLSKS-GNTPRPVEVHLPPSEELALNEL 276

Qy 542 VSLTCLVNGFVPSDIAVWESNGQ--PENNYKTPPVLD-SDG--SFPLYSKLTVDKSRW 596
    |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 277 VTLTCLVNGFSPKVDLVWMLQSGDELPREKYLTMASRQEPGQTTTFAVTSILRVAAEDW 336

Qy 597 QCGNVFSCVNHGALHNHYTOKSLSPG-----LQDDETC 632
    |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 337 KKGDPTFCMVGHEALPLAFTOKTIDRLAGKPTNVVSVVMAEVDGTC 383

```

RESULT 66

```

Q29028 Q29028 PRELIMINARY; PRT; 99 AA.
ID AC Q29028;
DT 01-NOV-1996 (TREMBLrel. 01. Created)
DT 01-NOV-1996 (TREMBLrel. 01. Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25. Last annotation update)
Db CD4, allele 2 (Fragment).
OS Sue scrofa (Fig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
NC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
NCBI_TaxID=9623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN#1183; TISSUE=Blood;
RX MEDLINE=93329116; PubMed=8335933;
RA "Guataveson K., Germana S., Sundt T.M., Sachs D.H., Leguern C.;
RT "Extensive allelic polymorphism in an exposed region of the
    minilature. . . . .";
RL J. Immunol. 151:1365-1370(1993).
DR EMBL; X65630; CAA46584.1; -.
DR PIR; I47132; S21462.
DR HSSP; P01730; 1CDY.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 99
SQ SEQUENCE 99 AA; 11390 MW; C229E5BEA28318F CRC64;

```

Query Match 8.2%; Score 280.5; DB 6; Length 99;
 Best Local Similarity 56.1%; Pred. No. 7.5e-14;
 Matches 55; Conservative 19; Mismatches 23; Indels 1; Gaps 1;

```

Qy 32 KKGDVETLCTAQQKSIQFMKNSNQIKILGNQGSFLTQGP-SKLANDRADSRSLMDQG 90
    |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 1 KAGDLAELPCHSSQKKNLPFSWKNSDQIKILSRHNLHMKASVTELSRLDSKKNMWDHG 60

Qy 91 NFPILIKLKTEDSPDYCEVEDQKEVQLLVFGITAN 128
    |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 61 SFPLIKLVLEVDGIYCEVEDKRIEQLLVFLRLTAS 98

```

RESULT 67

```

Q8WY24 Q8WY24 PRELIMINARY; PRT; 497 AA.
ID AC Q8WY24;
DT 01-MAR-2002 (TREMBLrel. 20. Created)
DT 01-MAR-2002 (TREMBLrel. 20. Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25. Last annotation update)
Db SN66 protein.
OC Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RA Zheng S., Shao X., Cao J., Geng L., Fang Y., Dong Q.;
RT "Identification and characterization of SN66, a Ig-like gene which is
    down-regulated in colorectal cancer.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF283666; AAL36987.1; -.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_MHC.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
SQ SEQUENCE 497 AA; 53665 MW; F24D08DPA5A6355 CRC64;

```

Query Match 8.2%; Score 279.5; DB 4; Length 497;
 Best Local Similarity 25.2%; Pred. No. 8.9e-13;
 Matches 118; Conservative 62; Mismatches 196; Indels 93; Gaps 21;

```

Qy 229 VEKLTSGGELMQAERASSKSWITFDLKNKEYSVK--RVT--QDPKLOMGKPLHLT 283
    |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 56 VROAPGGGLW-----MGWNNPQYGTNTFPAQKFGRLTSRDSINTAYVWLSLS 106

Qy 284 LPQALPOYAGSGNL-----TLLEAKTGKLNHOENVLVVNRATQLOKN- 325
    |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 107 TEDSAIFCARGNLRGGRGYWFPDWMGHTLVYSSAFTNPKVPLSLCSTQDPGNV 166

Qy 326 -LTCEVWG--PTSPKMLSLKENKEAKVSKREKPVVNLNPEAGMOCCLSDSGVLLS 382
    |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 167 VIACLVGFPFOEB--LSVTWSESGQGVYARNFP-----PSQASGDLVTTSSQLTLP 218

Qy 383 -----NKK-----VLPWTSTPPEPKSCDKHTKCPCPAPELLGGSVF 420
    |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 219 TQCLAGKSVTCHVGHYTNPSQDVTVPCEVPSTPPTP-----SPSTPPSPSCC-HPRLS 272

Qy 421 LFPKPKDPTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKRREQDYNSTYR 480
    |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 273 LHRPALED-LLLGSEANLTCTLTGL-RDASGVFTWTPSSGK--SAVQGPERRDLGGCVSVS 328

Qy 481 VSVLVTVLHODWLNKGEYKCKVSNKALPAPIEKTISAKAQPREPOVYTLPPSRDELTKN 540
    |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 329 VSSVLPQCAEPWNGKFTCTAAVPSKPTLTATLPKS-GNTPRPVEVHLPPSEELALN 387

Qy 541 Q-VSLTCLVNGFVPSDIAVWESNGQ--PENNYKTPPVLD-SDG--SFPLYSKLTVDKS 594
    |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 388 ELVTLTCLVNGFSPKVDLVWMLQSGDELPREKYLTMASRQEPGQTTTFAVTSILRVAAE 447

Qy 595 RMOQGNVFCVNHGALHNHYTOKSLSPG-----LQDDETC 632
    |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 448 DMKKGDPTFCMVGHEALPLAFTOKTIDRLAGKPTNVVSVVMAEVDGTC 496

```

RESULT 68

```

Q80Z17 Q80Z17 PRELIMINARY; PRT; 487 AA.
ID AC Q80Z17;
DT 01-JUN-2003 (TREMBLrel. 24. Created)
DT 01-JUN-2003 (TREMBLrel. 24. Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25. Last annotation update)
Db Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RA Strauberg R.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC049143; AAH49143.1; -.

```

DR InterPro; IPR003599; IG_1.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; Ig_3.
DR SMART; SM00409; IG_3.
DR SMART; SM00407; IGc1; 3.
DR SMART; SM00406; IGv; 1.
DR PROSITE; PS00835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
DR Hypothetical protein.
KW SEQUENCE 487 AA; 53019 MW; 31E2C893900A4D80 CRC64;

Query Match	Similarity	8.1%	Score 276.5	DB 11	Length 487
Best Local	Similarity	23.5%	Pred. No. 1.5e-12		
Matches 129	Conservative	80	Mismatches 216	Indels 123	Gaps 24
Qy	121	LVFGLTANSDDTHLLOGOSLITLTLES-----PPGSSPVSQCRSPGKNIOGGKTLSVSOL	174		
Db	1	MYLGINCIVFIYVLNGVQSEVLTLEAGGGLVQPGGSMKLSCAA-----SGFTF-----	48		
Qy	175	ELQDSGTCITCYLQÑ-ÖKKVEFKIDIVLAFQKASSIYVKKGEQVESEFFLAFIYEKLT	233		
Db	49	-----SNYMNWVRQSEPKGLEWVAERLMSNNYATHYAESVYGR-----FTISRDD	95		
Qy	234	GGSELMMQ-----AERASSSKSMITFDLKNKEVSVYKRVITQDPTKLQMGKLP	279		
Db	96	SKSSVYLLQMNLIARAEDTGIIYCTRRGYGDPMNY-FDVMGAGTIV-TVSEEP-----AREPT	149		
Qy	280	LH-LTLPOLP-----QYAGSGNLTLEAKTGKTHQEVNLYVMRATOLQKNLT	327		
Db	150	IYPLTFPOLASDPVILIGCLIHDFPPSGTMMNTM-CGSGKDIITVVFPPALASGGRYTS	208		
Qy	328	CEVMGPTSPKMLSLKLENKEAKVSKREKPVVNLVPEAGMOCILSDSQVLLESNIKYL	387		
Db	209	SQLTLP-----VECEGESVYCSVOHDSNPQVELNVNCP	243		
Qy	388	PTWSTPVEPKSCDKTHTCPCPAPELLGSPVFLPPEPKDITMTSRTEPVTGVVVDVSH	447		
Db	244	GICSP-----TTPPPSCQ-----PSLSLQRLALED-LLGSDASITCTLNGL--	286		
Qy	448	EDPE-VKNNMYVDGVEVHNAKTKPPEBOQNST--YRVSVLTVLHQDLNKEVYCKYS	503		
Db	287	RDPEGAVFTW-----EPSTGDAVQKKAQONCGGCYSVSVPGCAERNMSGASFCTVT	341		
Qy	504	NKALPAPLEKTIKSKAGQREPOVYTLPPSRBELTQÑQ-VSLTCLYKGFPGDIATWEMS	562		
Db	342	HPESDT-LTGITAKTYNTTFPOVHLPPPSSELTALNELVSLTCLVRAFPKEXVLVRLH	400		
Qy	563	NGQ---PENNYKTPPVLVSDSG-SFFLYSKLTVDSRMQGNVPSGSMYBHLNHHYTO	617		
Db	401	GNEELSPESYLVFEEPLKEPGEGATYLVTSVLRVSAIEWIKQDQYSQWVGHEALPNNFTQ	460		
Qy	618	KSLSLSPG 625			
Db	461	KTIDRLSG 468			
RESULT 69					
Q99KA4					
ID	Q99KA4	PRELIMINARY;	PRT;	487 AA.	
AC	Q99KA4				
DT	01-JUN-2001	(TREMBLrel. 17, Created)			
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)			
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)			
DB	Hypothetical protein.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
NCBI_TaxId	10090;				
NP	(1)				
SEQUENCE FROM N.A.					

RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004786; AA04786.1; -.
DR HSSP; P01810; 2FEJ.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KW Hypothetical protein.
SQ SEQUENCE 487 AA; 52554 MW; 7DC8E96DB33077B CRC64;

Query Match	8.0%	Score 274.5	DB 11	Length 487
Best Local Similarity	21.8%	Pred. No.2,1e-12		
Matches 142	Conservative 81	Mismatches 120	Indels 207	Gaps 28
QY	1	MNRGVPFPHLLLV- 	QALLPATQGNKVLGKGGTVELTCTAS--OKKSIOFHW--	53
DB	1	MNFGSLIFLVVLKVGQCEVQVLESQGLV---KPGSLSLSCAASGFTTSSVAMSVR		57
QY	54	---KNSNQIKILGNQGSFLLTKGSKLNDRADSRRLMDQGNFPLIKNLKIEDSDYIC		109
DB	58	QTPRKRLFWATISIDSGSY-TYYPDNVYGRFTTISDN-AKNNLYLQMSHLSESDTAMYYC		115
QY	110	EVEDQKEVQLVVLGTLTNSDPTHLLOGSLFLTTESPPGSSPV-QCRSPKGNIOGGKT		168
DB	116	---ARDMGSPFGGYSRFD---YWGQGITTYSSBARNPFTIPLPLPRA-----		159
QY	169	LSVSLQLDSDGWTCTVLQNKVEFKIDIVLAFOKASSIYVKKGEQVEFSPLAFT		228
DB	160	-----LSSDPVIL-----GCLINDYPP-----		176
QY	229	VEKLTGSGELMQAERASSSKSMITPDLKXKEVGVKRTQPKQMGKKLPLHLPLQAL		288
DB	177	-----SGTMNVYWGSKGQDITT-----VNFPLV		200
QY	289	PQ---YAGSGNLT/-ALEAKTGKLEHENVLVMBRATOLQKNLTGEVWGPTSPKMLSLKL		344
DB	201	ASGGGYTMSQULTLPAYEC-----		219
QY	345	ENKAKYSKRKEPVAWLVNPEKGMQCLISDSCQVLLSEINIKVLTWSTPVPKSCDKTHT		404
DB	220	-----PEGEYSKCSYQDSNVAQVEIDVVC-----SGPPPP-----		249
QY	405	CPPCPAPELGSPSVFLPPPKPQDTLMISRTPEYTCVVVDVSHEDPEKFMVYDVEVH		464
DB	250	CPPCP-PSC-HPSLSLDRPALPD-LLIGSDASLTCTLNGLRNPGAV-FTW-----EPS		299
QY	465	NAKTPREBOYNST---YRVVSVLTVLHQMVLNGKEYKCVKSNKALPAPIEKTISKAKQ		521
DB	300	TGXDAVQKKAQVNSCGCYSSVSVLPGCAERWNSGASFCTYTHPESDT-LGTIAKIVTN		358
QY	522	PREBOYTLPRSRBELTNQ-VSLTCLVKGYPEDIANEMESNQ---PENNYKTPPV		577
DB	359	TFPPQVHLPRPSSBELTALNELVSLTCLVRANPKEVLRKMLHGNBELSPESLTFEPLKE		418
QY	578	DSDG--SEFLYSKLTVDKSRWQGNVSGSVMEHALHNHYTOKSLSPG		625
DB	419	PGEGATTVLTVSVLRVSAETWKQDDQYSCVVGHEALPWNFTQKTIIDRLSG		468
RESULT 70				
ID	088650	PRELIMINARY;	PRT;	120 AA.
AC	088650;			
DT	01-NOV-1998 (Tremblrel. 08. Created)			
DT	01-NOV-1998 (Tremblrel. 08. Last sequence update)			
DT	01-OCT-2003 (Tremblrel. 25. Last annotation update)			
DE	T-cell surface glycoprotein CD4 (fragment).			
NN	CD4.			
OS	Marmota monax (Woodchuck).			

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Scurinae;
OC Marmota.
OX NCBI_TaxID=9995;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Periphereal blood;
RX MEDLINE=20094937; PubMed=10627561;
RA Guo J.T., Zhou H., Liu C., Aldrich C., Sapurelli J., Whitaker T.,
RA Barrera M.I., Mason W.S., Seeger C.;
RT "Apoptosis and regeneration of hepatocytes during recovery from
transient hepatitis virus infections.";
RL J. Virol. 74:1495-1505(2000).
DR EMBL; AF082497; AAC32621.1; -.
DR HSSP; P01730; 1WIO.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR000973; CD4_TcRg.
DR InterPro; IPR007110; IG-like.
DR PRINTS; PR00692; CD4TCANTIGEN.
FT NON_TER 1 1
FT 120 120
SQ SEQUENCE 120 AA; 13623 MW; 44967B512D520195 CRC64;

Query Match 8.0%; Score 273; DB 11; Length 120;
Best Local Similarity 64.0%; Pred. No. 3.7e-13;
Matches 57; Conservative 8; Mismatches 24; Indels 0; Gaps 0;
QY 308 LHQENLVVMAATQCKNLTCGEVWGPPTPKMLSLKLENKAKVSKRPVYVNLPEKGM 367
DB 1 LHQENLVVMAATVYHQNDLICEVGLGPTPKMLSLKLNQKAKVSKRPKIRVNPXKM 60
QY 368 WQCLLDSCQVLLSENIKVLPTWSTPVP 396
DB 61 WQCLLDSCQVLLSENIKVLPTWSTPVP 89

RESULT 71
Q8VCX4 PRELIMINARY; PRT; 489 AA.
AC Q8VCX4;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN IGH-VJ558 OR A1893585.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RA Strauberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; BC018322; AA18322.1; -.
DR MGD; MGI:96486; Igh-VJ558.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003066; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KW Hypothetical protein.
SQ SEQUENCE 489 AA; 53208 MW; CC85B1194DAFEF2C CRC64;

Query Match 8.0%; Score 272.5; DB 11; Length 489;
Best Local Similarity 24.5%; Pred. No. 3e-12;
Matches 135; Conservative 77; Mismatches 210; Indels 129; Gaps 27;
QY 118 VQLVFGITA--NSTHLLQ--GQSITLTLESPPSSSEVQCRSPRGKNIQGGKLSVSQL 174
DB 118 VQLVFGITA--NSTHLLQ--GQSITLTLESPPSSSEVQCRSPRGKNIQGGKLSVSQL 174

DB 6 VFLFLLSVTAGVHSKVLQOQSGAEV-----KPGASVLSCKA-----SGYTFSDYFI 53
QY 175 E--LQDSGT-----W-----TCTVLQNKQKVEFKIDIVLAFORAKSIVYKKEGEQVEFS 222
DB 54 HMKIKRSGQGLEMTGWNPPSGSISKPNK---FK-DATATLADKSSITTV----- 99
QY 223 FPLAFVTEKLTGSGELWMOER-----ASSKSWITFDLKNKEVSVK-RVTQDPKLOMG 275
DB 100 ----MDLSRLTSEDSAVYFCARHEDRGNVDGSLAMFYWGQGLTVTVAEAPRPPTI--- 152
QY 276 KKLPLHITLQALP-----QYAGGNLTALAEATGKLNQHVNLVMAATQLOK 324
DB 153 ----YPLTFPQALSSDPVITIGCLIHDPFSGTMVTV--GKSGKDTITVNEPPALASGARY 207
QY 325 NLTCGEVWGPPTPKMLSLKLENKAKVSKRPVYVNLPEKGMQCLLDSCQVLLSENI 384
DB 208 TMSQGLTLP-----VCPGEGSVKCGVQDSNVQGLNV 242
QY 385 KVLPTWSTPVEPKSCDKTHTCPPCAPELLGSPVFLPPPKPDTLMISRTPEVTCVVD 444
DB 243 NCPGICSP-----TTPPPSCQ-----FSLSLQRPALPD-LILGSDASITCTLNG 287
QY 445 VSHEDPE-VKFNWYVDGVEVHNAKTKRBEQVNST---YRVSVTLVTHQDMLNGKYYKC 500
DB 288 L--RDPGAVFTW-----EPSTGDAVQKRAVQNSCGSVSVLPCCARRWMSGASFCK 340
QY 501 KVSNAKLPADIEKTSISKAGQPREPVYTLPPSRDELTKNQ-VSLTCLVKGFPYSDIAVE 559
DB 341 TVTHPESDT-LTGIAVYTVNTTPPYQHLLPPSEELALNELVSLTCLVAEFNKVLYR 399
QY 560 WESNGQ---DENNYKTPPYVLDSDG--SFLYSKLTVDKSRMOQGNVFGSCVWHEALHNH 614
DB 400 WLHGNELSPESYLVFEPLEKPEGAGATTVLTVSVLRVSAEIMKQGDQYSQVNGHEALPMN 459
QY 615 YTKQSLSLSPG 625
DB 460 FTQKTDRLSLG 470

RESULT 72
Q90WB5 PRELIMINARY; PRT; 482 AA.
AC Q90WB5;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor.
OS Anas platyrhynchos (Domestic duck).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Anas.
OX NCBI_TaxID=8839;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RA Chan S.W.S., Middleton D.L., Lundqvist M., Warr G.W., Higgins D.A.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF378701; AAKS9279.1; -.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00409; IG; 3.
DR PROSITE; PS50835; IG_LIKE; 2.
KW Signal.
FT SIGNAL 1 29 POTENTIAL.
SQ SEQUENCE 482 AA; 54624 MW; BDA0B8E6172AD0B CRC64;

Query Match 8.0%; Score 271.5; DB 13; Length 482;
Best Local Similarity 25.0%; Pred. No. 3.5e-12;
Matches 111; Conservative 69; Mismatches 169; Indels 95; Gaps 19;
QY 8 RHLIVLQALLPAATGKNVLTGKGDYVELTCTASQKSIQPHMK-----NSNQIK 60
DB 12 RAVFVLLQGL--THMAHQOQIGVEKEVILNC---KKGDVTVTWREYDAGSSAIIQ 66

```

Qy      61 ILNQSSQFLTKGSGXKLNDAADSRSLMDQGNFLIKLKLIKEDSPYICVEDQKEVQL 120
Db      67 ILGK---IRKGAPMSDSSET-----NNSHGKLVSNLRISDAGTYICEGSDRNSISL 118
Qy      121 LVEGLTANSDPTHLLOQOSLTLT-LESPSPSSPS-----VQCSRPGKN 162
Db      119 HAVKLTITSSNGYLPEDDELYTMHKSPPSQRFSTLPFNSHNSKVLPVQLNETPQ--- 175
Qy      163 IQGKTLTSSVLELQDSGTWCTVLQNOKVKEFKI-DIVLAFQKAS-SIVYKKEGEYV 219
Db      176 ---KVALKVLQVLPQDSGTWICMHSDSPSINENISFNVKVLGFKEKTHLBERVAAVDSTV 232
Qy      220 EFSEPLAFVYBKLTG-----SGELMQARASSKSMITFDLKNKEVSYKRVTDPRQLQM 274
Db      233 TLSEWHLNF---RKIKGKEFFTGQLNMQEGNA-----ITTELDLPFNATADGELRETK--- 280
Qy      275 GKULPLHLTLPLQALP-----QVAGSGLTLALAEATGKXGHOVNLVWVARQ--- 321
Db      281 -KKSQALLERPEMKRSTVEVKIKHKIQLKHSEYTCQLLYNRRIYQSTTELVLVWQSANP 339
Qy      322 -----LQKULTCEVWGPPTSPKML-----SLTKENKEAKVSKREKPVVLNPEAGMW 368
Db      340 PGPRLPKGAEWTLTCQVSSPIRNVHLMERVNGTGMQDKSKSQS--ETKEVVKYTLAYGMW 397
Qy      369 QCLLSDSGQVULESN--IKVLPFW 390
Db      398 NCHLMEDNNMKSLNVTYVEAPTW 421

```

RESULT 73
Q99LA6
ID Q99LA6 PRELIMINARY; PRT; 484 AA.

DT 01-JUN-2001 (TReMBLrel. 17, Created)
 DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RA Straubeberg R.;
 RL Submitted (Feb.-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC003495; AA03495.1; -.
 DR PIR; F33932; F33932.
 DR HSSP; P01810; 2FBJ.
 DR InterPro; IPR007110; Ig_1like.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003596; Ig_v.
 DR Pfam; PF00047; Ig; 4.
 DR SMART; SMO0406; Igv; 1.
 DR PROSITE; PS50835; IG_LIKE; 4.
 DR PROSITE; PS00290; IG_MHC; 2.
 DR Hypothetical protein.
 QO SEQUENCE 484 AA; 52567 MW; 8EAE4AF9BCF582FA CRC64;

Query Match	8.0%;	Score 271.5;	DB 11;	Length 484;
Best Local Similarity	22.6%;	Pred. No. 3.5e-12;		
Matches 144;	Conservative 76;	Mismatches 217;	Indels 201;	Gaps 27

QY 13 VLQDALLPATQONKV-----VLKGGDTVSLTCTAQSKSIQF-HMKNSQIKI 61
Db 4 VMTLLFLMAAAGSIQAQIQLVQSGPELKKKPGETVXISCKAGYFTFDYSMMWKAQAFKG 63
QY 62 LGNGGSE-LITKGPSTKINDRADSRSH---WDQGNFLLITLKLIKIEDSDYICEVEDQKEE 117
Db 64 LKMGWNIETGSEVYADDFKGRFPAFSLTASTHLDINLNKNEPDATATFCARSDVDY 123
QY 118 VOLLVEGLTANSDTHLLQGSLLTLTLESPPGSSPVSQCRSPRGKNIQSGKTLVSQLEIQ 177

Db	121	IYAM-----DYMGGTSTVYSSSE-----ARNP-----IYYULPLP	154
Qy	178	DSGTCITCTYLOKQKVEFKIDIVLAFQKASSIYKKGEQVEFPFLAFTVEKLTGSGE	237
Db	155	PA-----LSSDPVIL-----GCLIHDPF-----SCTVMN	178
Qy	238	LWMQAEARASSSKWMTFPLKKNKEVSKVRQTDDPKLQMGKULPHLTLPOALPOYAGSGL	297
Db	179	VYV-----GKSGK-----DITVNEPPLASGR-----YTMSSQL	209
Qy	298	TL-ALEAKTGKLEQEVNLVVMKATOLQKHLCEVMGPTSPKMLSLKENKEAKVSKREK	356
Db	210	TLPAVACEPG-----ESVKCSVGHDSN	231
Qy	357	PVVVNLNPEAGMGMQCLISDSGVLLESNKVLPTWSTPEPKSCDKHTTCCPPCAPBELGG	416
Db	232	PVEELD-----VNCSG-----PTPPPP-----TIPSCQ-----	255
Qy	417	PSVFLPEPKPKDTLMISTRPEVTCVVVDVSHEDDEVKFNMYVDGVEVHNAAKTRPREOYN	476
Db	256	PSLISLGRPALPD-LILGSDASITCTLNGLRNPEGV-FTW-----EPSTGDAVQKKAVQ	308
Qy	477	ST---YRVVSVLTVLHODMLNGKEKCKVSKAKLPAPELKIETISKAKQPREPOVYTLPPS	533
Db	309	NSGCGSVSSVYPCGAERNMSGASFCKVTNPE-SGLTLGTIAKYVTNTPPVOHLLPPP	367
Qy	534	RDELITNQO-VSLTCLVKGFPSPDIAVEMESNGO---PENNYKTPPVLADSDG---SFFLYS	587
Db	368	SEELANELLSTLCTLVRAFNPKVAVYRVLHGNBELSPESYLVFEPPLBPBGCAATLYLVTS	427
Qy	588	KLTVDKSRMQGQNVFCSVMHEALHNHYTKSLSLSPG	625
Db	428	VLEVSALETWKQGDQVSCWVGHGALFPMNFTQKTIIDRLSG	465

RESULT 74
Q91Z07
ID Q91Z07
PRELIMINARY;
PRT; 486 AA

DT	01-DEC-2001	(TREMBLrel. 19, Created)
DT	01-DEC-2001	(TREMBLrel. 19, Last sequence update)
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)
DE	Hypothetical protein.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus	
OX	NCBI_TaxID=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Strausberg R.;	
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.	
DR	EMBL: BC010324; AAH0324.1; -	
DR	InterPro: IPR007110; IG_1like.	
DR	InterPro: IPR003006; IG_MHC.	
DR	InterPro: IPR003596; IG_V.	
DR	Pfam: PF00047; Ig; 4.	
DR	SMART: SM00406; IGV; 1.	
DR	PROSITE: PS50835; IG_LIKE; 4.	
DR	PROSITE: PS00290; IG_MHC; 2.	
KW	Hypothetical protein.	
QO	SEQUENCE 486 AA; 52682 MW; 4FE8F35125DA70B CRC64;	

Query Match	7.9%	Score 270.5;	DB 11;	Length 486;
Best Local Similarity	26.9%;	Pred. No. 4.2e-12;		
Matches 98;	Conservative 54;	Mismatches 143;	Indels 69;	Gaps 15

[illegible]

```

QY 391 STEVEPSCDKHTCPBPAPBLGGPSVFLPPPKDITLMISRPBXTCVVDVSHEDP 450
    243 SGPPPP-----CPCP-PSC-HPSLSLRPABED-LILGSAISLTCLTNLRPBG 290
Db 451 EYKFNMYDQGEVHNAKTKPREQYNST--YRVSVLTVLHQMIDMGKEYCKVYSNKAL 507
    291 AV-FTW-----EPSTCKDAIVOKKAQVNSCGCYSVSVSPGCAERNNSGASFKVTNPES 344
Db 508 PAPIEKTISKAKQPREPOVYTLPPSRDELTKNQ-VSLTLCYKGPSPDIAVEMESNGQ- 565
    345 DT-LTGITAKITNTTPPOVHLPPRSEELAINELVSLTCLYRAPNPEVLYRRLHGNEE 403
Db 566 --DENNYKTPPYLDSG--SFFLYSKLTVDKSRNQGNVFSQSYMEHALNNHYTQKSL 621
    404 LSPESYLVFEPLKEPPEGATTVLYLVTSVLRVSAETKQDOYSCMWGHEALPMNFTQKIID 463
Db 622 LSPG 625
    464 RLSG 467
Db

```

RESULT 75	.
Q91WT1	.
ID Q91WT1	PRELIMINARY;
Q91WT1	PRT; 481 AA

DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
OX NCBI_TaxId=10090;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE=Colon;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013490; AAH13490.1; -.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003586; IG_v.
DR Pfam; PF00047; Ig_4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS500290; IG_MHC; 2.
KM Hypothetical protein.
SQ SEQUENCE 481 AA; 52105 MW; 97DFE8D159463F65 CRC64;

Query Match	7.9%;	Score 270;	DB 11;	Length 481;
Best Local Similarity	22.0%;	Pred. No. 4.5e-12;		
Matches 139;	Conservative 69;	Mismatches 213;	Indels 212;	Gaps 25

QY	15	QLALLPAATGNVVLGKGGDTVELCTAS--QKXSIOFHWKMSNOIITLGNOC-----66
Dd	20	QVOLLGSPP-----LVFGASVKYSICKRSGTTFPSYYIH-----VKQRGGGLWIGM 69
QY	67	SPLTKGPSKLINDRADSRSLM---DQGNPELLIKULKIEDSDTYICEVEDOKEEVOLLVF 123
Dd	70	IYPGDNGTKYNKKFKKTLLTLADKSSSTAYMFLSLTSEDAVAFFC-----1151
QY	124	GLPANSPTHLLOQSLLTLESPPSPSPAVQGRSRRNNIQGKTLISVQLLEDDSGTW 183
Dd	116	-TTGGGMARDPYMGCGTTLTVSSEPARPEPT-----1433
QY	184	CTLVLONOKUVEKIDIVLAFOKASSIYVKKEGEVEFSFPPLAFTVEKLTGSGELMWQA 243
Dd	144	-----IYPLTFPOLSSDPVIILGCLIHDFP-----SGTMNVTV---1771
QY	244	RASSSKSMITFDLKKNVEKRVTDPRKLGOMKKULPHLTPALPQYAGSGNTTL-AL 302
Dd	178	GKSGK-----DITVNFEPLAASGR-----YTMSQQTLTPAVE 2111

```

QY 303 AKTGKHLGQVNLVWRARATQLOGNULCEVWGPSTPKMLMLSLKEKKAWSKREKPPVNL 362
Dh 212 CPEG----- : : : : :
QY 363 PEAGMWQCLSDSGVLLNESNIKVLPTWSTVEBPKSCDKHTCPCPAPILGSPSVFLF 422
Dh 234 VNC----- -IGICSP -TTPPPSCQ----PGLSLQ 258
QY 423 PPKPKDTLMISTPEVTCVVVDVSHEDPE-VKFNWYVDGVEFNHAKTTPKEBOQNTS--- 478
Dh 259 RPALED-LILGSDASITCTLNGL- -RODEGAVFTW-----EPSTGKDAVOGKKAQVNSCGC 310
QY 479 YRVSIVLTILHDDMLNGEKYEKCKSNKALPAPIEKTSKAKGQRFQVYTLPPSRDELT 538
Dh 311 YSVSVTLPCACARMMNSGASPKCTVTHPSBDT-LGTITAKTVNTFPQVHLLPPSEELA 369
QY 539 KQD-VSLTCLVYGFPSDIAVEMESNQ---PENNYKTTTPVLDSDG--SFLYLSKLTVD 592
Dh 370 LNELSVLTLCLVAENPKVILVRMLHGNSELSPESVLVEPLKEPGBGATTYLTVLSVLRS 429
QY 593 KSRWQGNVFSQVWHEALHNHYTQKSLSLPG 625
Dh 430 AEIWKQGDQYSCMVGHEALPMNFTQKIDRLSLG 462

```

RESULT 76	
Q8K172	
ID Q8K172	PRELIMINARY;
	PRT; 482 AA

DT	01-OCT-2002 (TREMBLERL. 22, Created)
DT	01-OCT-2002 (TREMBLERL. 22, Last sequence update)
DT	01-OCT-2003 (TREMBLERL. 25, Last annotation update)
DS	Similar to expressed sequence A1893585.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus
OX	NCBI_TaxId=10090;
RN	[1]
RN	SEQUENCE FROM N.A.
RC	TISSUE=Breast tumor;
RC	Strausberg R.;
RL	Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases
RL	EMBL; BC028249; AAH28249.1; -.
DR	PIR; F33932; F33932.
DR	PIR; PH1105; PH1105.
DR	PIR; PH1108; PH1108.
DR	PIR; PH1114; PH1114.
DR	PIR; PH1116; PH1116.
DR	PIR; PH1119; PH1119.
DR	PIR; PH1125; PH1125.
DR	PIR; PH1126; PH1126.
DR	PIR; PH1128; PH1128.
DR	PIR; PH1131; PH1131.
DR	PIR; PH1134; PH1134.
DR	PIR; PH1139; PH1139.
DR	PIR; PH1142; PH1142.
DR	PIR; PH1149; PH1149.
DR	PIR; PH1150; PH1150.
DR	PIR; PH1151; PH1151.
DR	PIR; PH1152; PH1152.
DR	PIR; PH1153; PH1153.
DR	InterPro; IPR007110; Ig-like.
DR	InterPro; IPR003587; Ig_C1.
DR	InterPro; IPR003006; Ig_MHC.
DR	InterPro; IPR003596; Ig_V.
DR	Pfam; PF00047; Ig; 4.
DR	SMART; SM00407; IG_C1; 3.
DR	SMART; SM00406; IGV; 1.
DR	PROSITE; PS00835; IG_LIKE; 4.
DR	PROSITE; PS00290; IG_MHC; 2.
DR	SEQUENCE 482 AA; 52121 MW; A06FF083E771D084 CRC64;

Query Match 7.9%; Score 270; DB 11; Length 482;
 Best Local Similarity 22.1%; Pred. No. 4.5e-12;
 Matches 136; Conservative 70; Mismatches 209; Indels 200; Gaps 24;

QY 30 LGKKDTELTCTAS--QKKSIOFW--KNSNQIKLQSGFLTKGPSKLNDRADSR 84
 DB 30 LKPGASVTLSCASAGYFTSYMMHWKORPGKGLWIGRIP--NSGGTKYNEKFKSKA 87
 QY 85 SLW---DQGNFLIKNLKIEDSDTYICEVEDQKEVQLVGLTANSTHLLQGSQTL 141
 DB 88 TLTVDKPSSTAYWQSLSTSEDSAVYYCTRE-----GDYAMDYWGQGSYV 133
 QY 142 TLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLQNKVFKDIYV 201
 DB 134 TVSSSEPARREPT-----IYF 147
 QY 202 LAFQKASIVYKKEGQVESPPLATVETKLTGSGELMQAERASSKSWITFDLKNKEV 261
 DB 148 LTFPQALSSDPVILGCLIHDFP-----SGTMMVTM-----GKSGK----- 183
 QY 262 SVKRYTQDPLQMGKPLHLTLPLQALPQYAGSGNLT-ALBAKTYGKLHGVNLVYMRAT 320
 DB 184 DITTVNFPPLASGGR-----YTWSQQLTLPAVCPREG----- 216
 QY 321 QLOKNLTCFVWGPTSPKMLSLKLENKAKVSKREKPVVWVLPBAGMQCLSDSGQVLL 380
 DB 217 -----BSVKCSVQHDSPFVVELVNNC----- 237
 QY 381 ESNIKVLPWSTPVEBKSCDHTHTCPCPAPBELLGSPVFLPFPKDTLMTSRPEVTC 440
 DB 238 -----FGICSP-TPRPPSCQ-----PILSLQRPALPD-LILGSDASITC 276
 QY 441 VVVDVSHDEPE-VKFWYVDGVEVHNAKTKPREEQYNSF---YRVSVTLVHODWLKNC 496
 DB 277 TLNGT--RDPEGAVFTW-----EPSTGKAVQKAVQSCGCVSVSLPGCAERKNSGA 329
 QY 497 EYKCKVSNALPAPIKETISKAKGQPREPOVYTLPPSRDELTKNQ-VSLTCLVKGFPYSD 555
 DB 330 SKKCTVTHESDT-LGTITAKTVNTFPQVHLPPSEBELANELVSLTCLVRAFPNPK 388
 QY 556 IAVWESNQ---PENNYKTTTPVLDSG--SFPLYSKLTVDKSRMQGQNVSCSYMHBA 610
 DB 389 VLVRLHGNBELSPESYLVEFPLKEPGBGATTVLVTSVLRVSAELIKQDQYSCWGHBA 448
 QY 611 LNNHTYQKSLSPG 625
 DB 449 LPMNFTOKTIDRLSG 463

RESULT 77
 Q9DCD9 PRELIMINARY; PRT; 426 AA.
 AC Q9DCD9; 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DE 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DB Adult male kidney cDNA, RIKEN full-length enriched library,
 clone:0610041A01, full insert sequence.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB1_TaxID=10090;
 RN NCBI_TaxID=10090;
 RP SEQUENCE FROM N.A.

RA Schriml L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Blake J., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Sakai K., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guetlich S., Hill D., Hofmann W., Hume D.A., Kamuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Saeki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,
 RA Wymshaw-Boris A., Yoshida K., Haegawa Y., Kawaji H., Kontsaki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection";
 RL Nature 409:685-690(2001).
 DR EMBL; AK002875; BAB22422.1; -
 DR PIR; F33932; F33932.
 DR HSSP; P01810; 2FBU.
 DR InterPro; IPR007110; IG-1like.
 DR InterPro; IPR003597; IG-cl.
 DR InterPro; IPR003006; IG_MHC.
 DR Pfam; PF00047; Ig_4.
 DR SMART; SM00407; IGC1_2.
 DR PROSITE; PS00835; IG_LIKE; 3.
 DR PROSITE; PS00290; IG_MHC; 2.
 DR PROSITE; PS00290; IG_MHC; 2.
 SQ SEQUENCE 426 AA; 45819 MW; 56E1275BA48F6F81 CRC64;

Query Match 7.9%; Score 269.5; DB 11; Length 426;
 Best Local Similarity 26.9%; Pred. No. 4.1e-12;
 Matches 98; Conservative 54; Mismatches 143; Indels 69; Gaps 15;

QY 282 LTLPOLP-----QYAGSGNLTALBAKTYGKLHGVNLVYMRATOLQKNLTEV 330
 DB 93 LTLPRALSSDPVILGCLIHDFPSTGMVTV--GKSGKDTTVNFPPLASGQGYTMSQL 151
 QY 331 WGPSTPKMLSLKLENKAKVSKREKPVVWVLPBAGMQCLSDSGQVLLSNIKVLPW 390
 DB 152 TLPA-----VPCPEGSYKCSGVQHDSPVQLDVC----- 182
 QY 391 STPVPKSCDHTHTCPCPAPBELLGSPVFLPFPKDTLMTSRPEVTCVVDVSHDP 450
 DB 183 SGPPP-----CPGCP-PSC--HPSLSLQRPALPD-LILGSDASLTLTNGLRNPEG 230
 QY 451 EVKFWYVDGVEVHNAKTKPREEQYNSF---YRVSVTLVHODWLKNEKCKVSNAL 507
 DB 231 AV-FTW-----EPSTGKAVQKAVQSCGCVSVSLPGCAERKNSASRKYTHES 284
 QY 508 PAPIETISKAKGQPREPOVYTLPPSRDELTKNQ-VSLTCLVKGFPYSDIAVWESNQ- 565
 DB 285 DT-LGTITAKTVNTFPQVHLPPSEBELANELVSLTCLVRAFPNPKVLRMLHGNB 343
 QY 566 --PENNYKTTTPVLDSG--SFPLYSKLTVDKSRMQGQNVSCSYMHBA LNNHTYQKSL 621
 DB 344 LSPESYLVEFPLKEPGBGATTVLVTSVLRVSAELIKQDQYSCWGHBA LPMNFTOKTID 403
 QY 622 LSPG 625
 DB 404 RLSG 407

RESULT 78
 Q91WT3 PRELIMINARY; PRT; 481 AA.
 AC Q91WT3; 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DE 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DB Hypothetical protein.
 GN IGH-VJ558 OR A1893585.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB1_TaxID=10090;
 RN NCBI_TaxID=10090;
 RP SEQUENCE FROM N.A.

RC TISSUE=Colon;
 RA Strausberg R.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC013488; AAH13488.1; -
 DR MGI: 96486; Igh-VJ558.
 DR InterPro: IPR007110; Ig-like.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003596; Ig_V.
 DR Pfam: PF00047; Ig; 4.
 DR SMART: SM00406; IGV; 1.
 DR PROSITE: PS50835; IG_LIKE; 4.
 DR PROSITE: PS00290; IG_MHC; 2.
 KM Hypothetical protein.
 SQ SEQUENCE 481 AA; 52022 MW; 4EEB5C253038B718 CRC64;

Query Match 7.8%; Score 266; DB 11; Length 481;
 Best Local Similarity 23.2%; Pred. No. 9.1e-12;
 Matches 121; Conservative 75; Mismatches 187; Indels 138; Gaps 21;

QY 182 WTCTVL-----QNKQKVEFKIDIVLAFQKASIVYKKEGQVEFPFLA---FT- 228
 DB 3 WSCILFLVATATGVNSQVQLQ-----QPGAEIV--RPGASVLSCKTSQYFTD 50
 QY 229 -----VEKLTSGGELMWA-----ERASSKSWITFDLKNKEVS 262
 DB 51 YMMNWKQRPQGLEWIGAIDPDSYTSYNQKFKGTTLVDTSSSAYMLSLTSEDSA 110
 QY 263 VKRVTQDPKLQWG-----KKLPLHLTLPOLP-----QYAGS 294
 DB 111 VYFCARGRDSGGYWGQTTLTVSSEPARERTYPLFPQALSDPVIIGLIHDYFPS 170
 QY 295 GNLTALAEKTKGLHQNVLVVRATOLQKNLTCEVWGTPSPKMLSLKENKAKVSKR 354
 DB 171 GTMNVTW-GKSGKIDITVNFPPALASGGRTWSQTLPLA----- 209
 QY 355 EKPWVNLPEAGMOCCLSDSGVLESNIKVLPWSTPVEPKSCDKHTGCPAPELL 414
 DB 210 -----VECPGEGSVCSVQHDSPVQELNVNCPGICSP-----TTPPPSCQ-- 252
 QY 415 GSPVFLPPPKDITLMSRTEPVTGVVDVSHEDPE-VKFNWYDGVENNAKTPREE 473
 DB 253 --PSLSIORPALED-LLGSDASITCTLNGL--RDEGAVFTW-----EPSTGDAVQKK 302
 QY 474 QYNST--YRVVSVLTVLHQMVLNGKEYCKVSNKALPAPIEKTISKAKGPREQVYTL 530
 DB 303 AVQNSCGCYSVSVLPGCAERNNSGASFCKVTJHESDT-LTGTLAKTVVNTFPPQVHL 361
 QY 531 PPSRELTKNQ-VSLTCLVKGFPDIAVWESNQ---PENNYKTTTPVLDSDG--SFF 584
 DB 362 PPSRELTKNELVSLTCLVRAFPKXVLRWLDHGNELSPESYLVFEPLKEPGEATTYL 421
 QY 585 LYSKLTVDKSRWQGNVSCSYMHBALHNHYTKSLSPG 625
 DB 422 VTSVLRSAEIKWQGDQYSCWVGHEALPMNFTQKTIIDRLSG 462

RESULT 79
 Q91X92 PRELIMINARY; PRT; 482 AA.
 AC Q91X92;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon;
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: BC01181; AAH1181.1; -
 DR PIR: F33932; F33932.
 DR InterPro: IPR007110; Ig-like.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003596; Ig_V.
 DR Pfam: PF00047; Ig; 4.
 DR SMART: SM00406; IGV; 1.
 DR PROSITE: PS50835; IG_LIKE; 4.
 DR PROSITE: PS00290; IG_MHC; 2.
 KM Hypothetical protein.
 SQ SEQUENCE 482 AA; 51865 MW; 312E01F9C1BC7F3C CRC64;

Query Match 7.8%; Score 266; DB 11; Length 482;
 Best Local Similarity 22.2%; Pred. No. 9.1e-12;
 Matches 142; Conservative 71; Mismatches 224; Indels 202; Gaps 26;

QY 10 LLLVLQALLPAATQGNKVLGKK-----DIVELTCTAS--QKSIQFMWNSQIK 60
 DB 4 LALLCLVTPPSCAL-SQVQLKSGPDLVAPSQSLITCTVSGFALTSVAISWVROPCK 62
 QY 61 ILNGSGFLTKGSPSKLNDRASSRL-WDQGNFPLIKNLKIEDSDT--YICEVEDQKEE 117
 DB 63 GLEWLGIVTWGVTNNNSALKSLISKNSKQGVFLKNNSLQNDIARYYC-ARDNYYE 121
 QY 118 VQLVFGELTANSDTHLQGSLLTLESPPGSSPSVQCRSPKNGIKGKTLVSQLELQ 177
 DB 122 GAMDYV-----GQGTSTVSSSEPARERT----- 144
 QY 178 DSGTWCTVLQONQKVEFKIDIVLAFQKASIVYKKEGQVEFPFLAFTVEKLTSGE 237
 DB 145 -----IYPLTFPOLSSDPVIIGLIHDYFP-----SGTMN 175
 QY 238 LMMQAEERASSKSWITFDLKNKEVSVKRVTQDPKLQWKKLPLHLPOLPYAGSGNL 297
 DB 176 VTW-----GKSGK-----DITVNFPPALASGGR-----YTMSQL 206
 QY 298 TL-ALEAKTKGLHQNVLVVRATOLQKNLTCEVWGTPSPKMLSLKENKAKVSKREK 356
 DB 207 TLPVVECPG-----EVCVSVQHDSP----- 228
 QY 357 PWWNLNPEAGMOCCLSDSGVLESNIKVLPWSTPVEPKSCDKHTGCPAPELLGG 416
 DB 229 PVQELNVN-----GICSP--TTPPPSCQ----- 253
 QY 417 PSVFLPPPKDITLMSRTEPVTGVVDVSHEDPE-VKFNWYDGVENNAKTPREEQY 475
 DB 254 PSLSIORPALED-LLGSDASITCTLNGL--RDEGAVFTW-----EPSTGDAVQKAV 305
 QY 476 NST--YRVVSVLTVLHQMVLNGKEYCKVSNKALPAPIEKTISKAKGPREQVYTLPP 532
 DB 306 QNSCGCYSVSVLPGCAERNNSGASFCKVTJHESDT-LTGTLAKTVVNTFPPQVHLPP 364
 QY 533 SRDELTKNQ-VSLTCLVKGFPDIAVWESNQ---PENNYKTTTPVLDSDG--SFFLY 586
 DB 365 PSELTKNELVSLTCLVRAFPKXVLRWLDHGNELSPESYLVFEPLKEPGEATTYLV 424
 QY 587 SKLTVDKSRWQGNVSCSYMHBALHNHYTKSLSPG 625
 DB 425 SVLRVSAEIKWQGDQYSCWVGHEALPMNFTQKTIIDRLSG 463

RESULT 80
 Q91WP5 PRELIMINARY; PRT; 479 AA.
 AC Q91WP5;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;

RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon;
 RA Strausberg R.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC013656; AAH13656.1; -
 DR InterPro; IPR007110; IG_1like.
 DR InterPro; IPR003006; IG_MHC.
 DR InterPro; IPR003596; IG_V.
 DR Pfam; PF00047; Ig_4.
 DR SMART; SM00406; IgV_1.
 DR PROSITE; PS50835; IG_LIKE; 4.
 DR PROSITE; PS00290; IG_MHC; 2.
 KW Hypothetical protein.
 SQ SEQUENCE 479 AA; 51603 MW; ECB2D0877748584F CRC64;

Query Match 7.8%; Score 265.5; DB 11; Length 479;
 Best Local Similarity 24.7%; Pred. No. 9.8e-12;
 Matches 134; Conservative 78; Mismatches 207; Indels 123; Gaps 27;

QY 123 FGLTNSDTHLQGGSLTLT-ES-----PGSSPVSQCRSRGKNIQGGKTLVSQLEL 176
 DB 3 FGLTLIFLVLTGKGVQVLESVGGGLVPGSLKVSQCA-----SGLTFSSNYAM-- 53
 QY 177 QDSGTWCTVLQN-QKKEVEFKIDIVLAFQKASSIYKKEGQVEFPFLATVEKLTGS 235
 DB 54 ----SW---VROSPKRLKM-----VAALNSNGCNTYSD-----TMKGRFTISRDNK 95
 QY 236 GELIMQAERASSSKSNIT-----FDLKNKEVSVKRTQDPKLGKGLPLH-LTL 284
 DB 96 STLYLQ-----MSSLRSEDIAFYCCVRGYPDVWAGATAV--TVSESP-----AREPTIYPLTF 147
 QY 285 PQAALP-----QYAGSGNLTALBAKTKLHQEVNLVMPATOLQKNLTGEVMP 333
 DB 148 PQLSSDPVYIGCLLHIDYFSPSGTMMVNTW-GKSGKDIITTVNFPALASGGRYTMSSQLTLP 206
 QY 334 TSPKMLSLKLENKAKVSKREKPVVNLNPEAGMMQCLSDSGQVLLSNIKVLPTWSTP 393
 DB 207 A-----VECPGESVYKCSVQHDSNPFQELNVNCPGICSP 241
 QY 394 VEPKSCDKHTNCPCPAPALLGSPVFLFPPPKDITLMISRTPEVTCVVVDVSHEDP- V 452
 DB 242 -----TTPPPSCQ-----PSLSLQRPALD-LILGSDASITCTLNG- -RDPEGA 284
 QY 453 KFNMYVDGEVHNNAKTKPREEOYNST---YRVVSVLTVLHQDLNGEKYKCKVSKALPA 509
 DB 285 VFTW-----EPSTGDAVQKAVQNSCCGYSSVLPFGCAERWNSGASFKCTVTHESDP 339
 QY 510 PIEKTISKAKQPREPOVYTLPPSRDELTKNQ-VSLTCLVKGFPYSDIAVEMESNGQ--- 565
 DB 340 -LTGTLIAKTVNTFPQVHLIPRSEELALNELVSLTCLVRAFNPKVAVLRLHGNBELS 398
 QY 566 PENNYKTPPVLDSDG--SFPLYSKLTVDKSRMQGNVSCSMHEALHNHYTOKSLSLIS 623
 DB 399 PESYLVEFPELKEPEGATTVLTVSVLRSAELIWKQDDQYSCMVGHEALPMNFTQXTIDRL 458
 QY 624 PG 625
 DB 459 SG 460

RESULT 81
 Q8VEAO PRELIMINARY; PRT; 484 AA.
 AC Q8VEAO;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Hypothetical protein
 GN IGH-VJ358 OR A1893585.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon;
 RA Strausberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC019425; AAH19425.1; -
 DR MGD; MGI:96486; Igh-VJ558.
 DR InterPro; IPR007110; IG_1like.
 DR InterPro; IPR003006; IG_MHC.
 DR InterPro; IPR003596; IG_V.
 DR Pfam; PF00047; Ig_4.
 DR SMART; SM00406; IgV_1.
 DR PROSITE; PS50835; IG_LIKE; 4.
 DR PROSITE; PS00290; IG_MHC; 2.
 KW Hypothetical protein.
 SQ SEQUENCE 484 AA; 52859 MW; F2522DF5ED6288A6 CRC64;

Query Match 7.7%; Score 264; DB 11; Length 484;
 Best Local Similarity 26.6%; Pred. No. 1.3e-11;
 Matches 97; Conservative 53; Mismatches 145; Indels 70; Gaps 14;

QY 282 LTLPOALP-----QYAGSGNLTALBAKTKLHQEVNLVMPATOLQKNLTGEV 330
 DB 150 LTFPQLSSDPVYIGCLLHIDYFSPSGTMMVNTW-GKSGKDIITTVNFPALASGGRYTMSSQL 208
 QY 331 WGPSPKMLSLKLENKAKVSKREKPVVNLNPEAGMMQCLSDSGQVLLSNIKVLPTW 390
 DB 209 TLPA-----VECPGESVYKCSVQHDSNPFQELNVNCPGIC 243
 QY 391 STVPEKSCDKHTNCPCPAPALLGSPVFLFPPPKDITLMISRTPEVTCVVVDVSHEDP 450
 DB 244 SP-----TTPPPSCQ-----PSLSLQRPALD-LILGSDASITCTLNG- -RD 286
 QY 451 E-VKFNMYVDGEVHNNAKTKPREEOYNST---YRVVSVLTVLHQDLNGEKYKCKVSKA 506
 DB 287 EGAVFTW-----EPSTGDAVQKAVQNSCCGYSSVLPFGCAERWNSGASFKCTVTHPE 341
 QY 507 LPAPIEKTISKAKQPREPOVYTLPPSRDELTKNQ-VSLTCLVKGFPYSDIAVEMESNGQ 565
 DB 342 SDF-LTGTLIAKTVNTFPQVHLIPRSEELALNELVSLTCLVRAFNPKVAVLRLHGN 400
 QY 566 ---PENNYKTPPVLDSDG--SFPLYSKLTVDKSRMQGNVSCSMHEALHNHYTOKSL 620
 DB 401 ELSPESYLVEFPELKEPEGATTVLTVSVLRSAELIWKQDDQYSCMVGHEALPMNFTQXTI 460
 QY 621 SLSPG 625
 DB 461 DRLSG 465

RESULT 82
 Q8KOF2 PRELIMINARY; PRT; 488 AA.
 AC Q8KOF2;
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Similar to expressed sequence A1893585.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon;
 RA Strausberg R.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC011703; AAH11703.1; -
 DR InterPro; IPR007110; IG_1like.
 DR InterPro; IPR003597; IG_c1.
 DR InterPro; IPR003006; IG_MHC.
 DR InterPro; IPR003596; IG_V.
 DR Pfam; PF00047; Ig_4.

DR SMART; SM00407; IGcl; 3.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
SQ SEQUENCE 488 AA; 53127 MW; 0E3B156E15733F0 CRC64;

Query Match 7.7%; Score 264; DB 11; Length 488;
Best Local Similarity 26.6%; Pred. No. 1.3e-11;
Matches 97; Conservative 53; Mismatches 145; Indels 70; Gaps 14;

QY 282 LTLQALP-----QYAGSGNLTLALEAKTGKLGHEVNLVWMRAATOLQKNTCEV 330
DB 154 LTFPQALSSDPVLIIGCLIHDPSPGTMNVTW-GKSGDITTVNFPALASGGRYTMSSQL 212
QY 331 WGPSTPKMLSLKLENKAKVSKREKPYVNLNPEAGMOCLLSDSGVLLSNIKVLPTW 390
DB 213 TLPA-----VECPGESVKSGVGHDSNPVGLANNCGIC 247
QY 391 STPEPKSCDKHTHTCPCPAPPELLGSPVFLPPPKKDTLMIISRTPEVTCVVVDVSHEDP 450
DB 248 SPP-----TTPPPPSQC---PSLSLQRPALSD-LILGSDASITCTLNGI--RDP 290
QY 451 E-VKFNWYVDGVNHNATKTRREQYNST---YRVSVTLVHQMVLNKGKCYKSNKA 506
DB 291 EGAVFTW-----EPSTGKDAVQKKAQVNSCGCYSVSVLPGCAERKNSGASFKCTVTHPE 345
QY 507 LPAPLEKTIKAKGQPREPQVYTLPPSRDELTKNQ-VSLTCLVKGFYPSDIAVEMESNQ 565
DB 346 SDF-LTGITAKVTNTPFPVHLLPPSEELANELLSTLCVRAFPKRVLRMLHNE 404
QY 566 ---PENNYKTPPVLDSDG--SFPLYSKLTVDKSRMOQGNVFSQVMEALHNHYTQSL 620
DB 405 ELSPESYLVEPFLKEPGEATTVLTVLSVSAEIMWQGDQYSCMVGEALPMNFTQXTI 464
QY 621 SLSPG 625
DB 465 DRLSG 469

RESULT 83

QY 091XEL PRELIMINARY; PRT; 480 AA.
AC 091XEL; 1.
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RA Strauberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010798; AAH10798.1; -
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 480 AA; 51936 MW; 20B9234EEF2B41ED CRC64;

Query Match 7.6%; Score 261; DB 11; Length 480;
Best Local Similarity 26.3%; Pred. No. 2.2e-11;
Matches 96; Conservative 54; Mismatches 145; Indels 70; Gaps 14;

QY 282 LTLQALP-----QYAGSGNLTLALEAKTGKLGHEVNLVWMRAATOLQKNTCEV 330

DB 146 LTFPQALSSDPVLIIGCLIHDPSPGTMNVTW-GKSGDITTVNFPALASGGRYTMSSQL 204
QY 331 WGPSTPKMLSLKLENKAKVSKREKPYVNLNPEAGMOCLLSDSGVLLSNIKVLPTW 390
DB 205 TLPA-----VECPGESVKSGVGHDSNPVGLANNCGIC 239
QY 391 STPEPKSCDKHTHTCPCPAPPELLGSPVFLPPPKKDTLMIISRTPEVTCVVVDVSHEDP 450
DB 240 SPP-----TTPPPPSQC---PSLSLQRPALSD-LILGSDASITCTLNGI--RDP 282
QY 451 E-VKFNWYVDGVNHNATKTRREQYNST---YRVSVTLVHQMVLNKGKCYKSNKA 506
DB 283 EGAVFTW-----EPSTGKDAVQKKAQVNSCGCYSVSVLPGCAERKNSGASFKCTVTHPE 337
QY 507 LPAPLEKTIKAKGQPREPQVYTLPPSRDELTKNQ-VSLTCLVKGFYPSDIAVEMESNQ 565
DB 338 SDF-LTGITAKVTNTPFPVHLLPPSEELANELLSTLCVRAFPKRVLRMLHNE 396
QY 566 ---PENNYKTPPVLDSDG--SFPLYSKLTVDKSRMOQGNVFSQVMEALHNHYTQSL 620
DB 397 ELSPESYLVEPFLKEPGEATTVLTVLSVSAEIMWQGDQYSCMVGEALPMNFTQXTI 456
QY 621 SLSPG 625
DB 457 DRLSG 461

RESULT 84

QY 08VCV5 PRELIMINARY; PRT; 481 AA.
AC 08VCV5; 1.
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast; tumor;
RA Strauberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC018455; AAH18455.1; -
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KW Hypothetical protein.
SQ SEQUENCE 481 AA; 52326 MW; 52B44C5826807143 CRC64;

Query Match 7.5%; Score 256.5; DB 11; Length 481;
Best Local Similarity 22.9%; Pred. No. 4.8e-11;
Matches 124; Conservative 82; Mismatches 215; Indels 121; Gaps 25;

QY 120 LLVFGLTANSDTHLQOGSITLTLESB---PSSPSVQCRSPRGKNI-----QGG 166
DB 6 IFLFLSLGTTGVH---SEIQLOQSGPELVRPSTSVVSCSKAGYSFIDVNIYVWKOSHG 61
QY 167 KTLG-VSQLELDQSGTMTCTVLDNQKVKERKIIIVLAFAKASISYVKKGEQVEFEPFL 225
DB 62 KSLBEWIGYIDPNYNGS---SYNQK--PK-GKATLTVDKSSNTAF-----99
QY 226 AFVTEKLTSGGEIWMQERASSKSWITPDKNKEVSK-RVTDPRKLOMGKLPDLHLTL 284
DB 100 -MYLANLITSDSAFYFCAKRWYG-AMPAFWGGTIVTVVSAESARNPTI-----YLLTL 150
QY 285 PQALP-----QYAGSGNLTLALEAKTGKLGHEVNLVWMRAATOLQKNTCEVWGP 333

```

Db      151 PALSSDPYICGLIHDYFPSCGTMTWTM-GKSGDITTVNFPALASGGRYTMSSQLTLP 209
Qy      334 TSPKMLSLKLENKEAKVSKREKPVVNLPEAGMOCLLSDSGVLLSENIKVL-PTWST 392
Db      210 A-----VECPGEBSVYCSVQHDSNPQELDVNCSGPTPP 244
Qy      393 PVEPKSCDHTHTCPPEAPBELGSPSVFLPPPKKDTLMIISRTPEVTCVVVVSHEDEPV 452
Db      245 PI-----TIPSCQ-----PSLSIQRPALPD-LILGSDASTITCLNLRNPEGAV 287
Qy      453 KKNWYVDGVVHNAKTKPREEQYNST---YRVSVLTVLHODMLNGKEYKCVSNKALPA 509
Db      288 -FTW-----EPSTGKDAVQKAVQNSCGCYSVSVLPGCAERNMNSGASFCTVTHPE-SG 340
Qy      510 PLEKTISSKAGQPREQVYVTLPPSRDELTKNQ-VSLTCLVKGFYPSDIAVEMESNQ--- 565
Db      341 TLTGITLAKVTNTFPQVHLPPPESEELINELSLTCLVRAFNPKEVLRVWLHGNBELS 400
Qy      566 PENNYKTPPVLDSDG--SFPLYSKLTVDKSRMOQGVNFSQSVMEALHNHYTKSLSL 623
Db      401 PSSYLVEFLKEBEGATTVLTYSVLRSATWKQGDQYSCMVGHFALPMNFTOKTIDRL 460
Qy      624 PG 625
Db      461 SG 462

RESULT 85
Q7TMK4 PRELIMINARY; PRT; 479 AA.
AC 07TMK4;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DE Hypochemical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6d; TISSUE=Breast tumor;
RX MEDLINE=2388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Scheffen C.M., Schuler G.D.,
RA Altchul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uebachs T.B., Toshlyuk S., Carninci P., Frange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullighy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kerteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skala U., Smalls D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6d; TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (Aug-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC055905; AAH55905.1; -
KW Hypochemical protein.
SQ SEQUENCE 479 AA; 52209 MW; 532DED9D46D0AED CRC64;

```

```

Query Match 7.4%; Score 252; DB 11; Length 479;
Best Local Similarity 23.8%; Pred. No. 1e-10;
Matches 110; Conservative 62; Mismatches 167; Indels 124; Gaps 21.

Qy      236 GELW--WQAEARSSKSWITFDLNKEV-----SVK-RVT---QDPK---LQNGKK 277
Db      49 GDVWNNWVROSPEKLEWVS-QIRNKPNYETVYSDSVKGFITISRDSSKSGVYLQNNML 107
Qy      278 LP-----LHVLTPQLP-----QYA 292
Db      108 RPEMDGIYTCVEGMDYGRIGISVTVSSSARNPITYLTPPALSSDPYIICLIHDYF 167
Qy      293 GSGNLTLLAEKTKLHQEVLVWRATQLQNLTCVEWGTSPKMLSLKLENKEAKVS 352
Db      168 PSGTMTNVTM-GKSGDITTVNFPALASGGRYTMSSQLTLP----- 208
Qy      353 KREKPVVNLPEAGMOCLLSDSGVLLSENIKVL-PTWSTPVEPKSCDHTHTCPPEAP 411
Db      209 -----VECPGEBSVYCSVQHDSNPQELDVNCSGPTPPPI-----TIPSCQ-- 250
Qy      412 ELIGSPSVFLPPPKKDTLMIISRTPEVTCVVVVSHEDEPVKFMWYVDGVVHNAKTKPR 471
Db      251 -----PSLSIQRPALPD-LILGSDASTITCLNLRNPEGAV-FTW-----EPSTGKDAVQ 298
Qy      472 EEQYNST---YRVSVLTVLHODMLNGKEYKCVSNKALPAPIEKTISKAGQPREQVY 528
Db      299 KKAIVNSCGCYSVSVLPGCAERNMNSGASFCTVTHPE-SGLTGTITLAKVTNTFPQVH 357
Qy      529 TLPPSRDELTKNQ-VSLTCLVKGFYPSDIAVEMESNQ---PENNYKTPPVLDSDG--S 582
Db      358 LLPPESEELINELSLTCLVRAFNPKEVLRVWLHGNBELSPESYLVEFLKEBEGATTV 417
Qy      583 FFLYSKLTVDKSRMOQGVNFSQSVMEALHNHYTKSLSLSPG 625
Db      418 YLVTSVLRSATWKQGDQYSCMVGHFALPMNFTOKTIDRLSG 460

RESULT 86
Q8K0Z4 PRELIMINARY; PRT; 480 AA.
ID Q8K0Z4
AC Q8K0Z4;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Similar to expressed sequence A1893585.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (May-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC029188; AAH29188.1; -
DR InterPro; IPR003599; IG_1-like.
DR InterPro; IPR007110; IG_1-like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG_4.
DR SMART; SM00409; IG_3.
DR SMART; SM00407; IG_3.
DR SMART; SM00406; IG_1.
DR PROSITE; PS00835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
SQ SEQUENCE 480 AA; 51645 MW; 8690A63C669CDBED CRC64;

```

```

Query Match 7.4%; Score 251.5; DB 11; Length 480;
Best Local Similarity 25.5%; Pred. No. 1.1e-10;
Matches 93; Conservative 56; Mismatches 145; Indels 71; Gaps 14;

Qy      282 LTLPPALP-----QVAGSGNLTLLAEKTKLHQEVLVWRATQLQNLTCVEV 330

```

```

Db      147 LTLPLALSSDPVITIGCLIHDFPSGTMVW-GRSGKDIITVNFPPALASGRYTMSSQL 205
Qy      331 WGPSPKMLSLKLENKAKVSKREKFWVNLNPEAGMOCILSSGOVLLESNTKVL-PT 389
Db      206 TLPA-----VCEPSESVKCSGVQHDNSNFGQLDVNCSGPT 240
Qy      390 WSTPVEPKCDKTHTCPPCPAPABELLGPSVFLPPKPKDTLMI SRTPEVTELVVDVSHED 449
Db      241 PPPP-----TIPSCQ-----PSLSLQRALED-LILGSDASITTLGLNRP 283
Qy      450 PEVKFNWYDGVENVNAKTKPREEOYNST---YRVVSVLTVLHODMLNGEKCKVSNKA 506
Db      284 GAV-FTW-----EPSTGKDAVQKKAQVNSCCYSVSVLPGCAERNMSGASFCTVTHPE 337
Qy      507 LPAPLEKTIKAKQGPPEPOVYTLPPSRDELTKNQ-VSLTCLVKGFPYSDIAVEMESNQ 565
Db      338 -SGTLTGTLAKVTNTPPPQVHLLPPSEBELNELSLTCLVRAFNPKVFLVRLHNGE 396
Qy      566 ---PENNYKTPPVLDSDG--SFPLYSKLTVDKSRMOGQNVFSCSVNHEALHNHYTKSL 620
Db      397 ELSPESVLYVEPLKEPREGATTYLVTSVLRVSAETWKQGDQYSCVNGHEALPNMFTQKTI 456
Qy      621 SLSPG 625
Db      457 DRLSG 461

```

RESULT 87

```

Q99M22 099M22 PRELIMINARY; PRT; 479 AA.
AC Q99M22;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC002091; AA02091.1; -.
DR HSSP; P01810; 2FBJ.
DR InterPro; IPR007110; IG_1-like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG_4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KW Hypothetical protein.
SQ SEQUENCE 479 AA; 51992 MW; 768E39A138918892 CRC64;

```

Query Match 7.3%; Score 250.5; DB 11; Length 479;
 Best Local Similarity 25.5%; Pred. No. 1,4e-10;
 Matches 93; Conservative 56; Mismatches 145; Indels 71; Gaps 14;

```

Qy      282 LTLPLALP-----QYAGSGNLTLEAKTKLHQENVNLVVMRATQLOKRLTCEV 330
Db      146 LTLPLALSSDPVITIGCLIHDFPSGTMVW-GRSGKDIITVNFPPALASGRYTMSSQL 204
Qy      331 WGPSPKMLSLKLENKAKVSKREKFWVNLNPEAGMOCILSSGOVLLESNTKVL-PT 389
Db      205 TLPA-----VCEPSESVKCSGVQHDNSNFGQLDVNCSGPT 239
Qy      390 WSTPVEPKCDKTHTCPPCPAPABELLGPSVFLPPKPKDTLMI SRTPEVTELVVDVSHED 449
Db      240 PPPP-----TIPSCQ-----PSLSLQRALED-LILGSDASITTLGLNRP 282
Qy      450 PEVKFNWYDGVENVNAKTKPREEOYNST---YRVVSVLTVLHODMLNGEKCKVSNKA 506

```

```

Db      283 GAV-FTW-----EPSTGKDAVQKKAQVNSCCYSVSVLPGCAERNMSGASFCTVTHPE 336
Qy      507 LPAPLEKTIKAKQGPPEPOVYTLPPSRDELTKNQ-VSLTCLVKGFPYSDIAVEMESNQ 565
Db      337 -SGTLTGTLAKVTNTPPPQVHLLPPSEBELNELSLTCLVRAFNPKVFLVRLHNGE 395
Qy      566 ---PENNYKTPPVLDSDG--SFPLYSKLTVDKSRMOGQNVFSCSVNHEALHNHYTKSL 620
Db      396 ELSPESVLYVEPLKEPREGATTYLVTSVLRVSAETWKQGDQYSCVNGHEALPNMFTQKTI 455
Qy      621 SLSPG 625
Db      456 DRLSG 460

```

RESULT 88

```

Q99M67 099M67 PRELIMINARY; PRT; 487 AA.
AC Q99M67;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE CD4 protein precursor.
GN CD4.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=H.B2; TISSUE=Thymus;
RX MEDLINE=9218434; PubMed=10201936;
RA Koskinen R., Lammimäki U., Tregaskes C.A., Salomonson J., Young J.R.,
RA Vainio O.;
RT "Cloning and modeling of the first nonmammalian CD4.";
RL J. Immunol. 162:4115-4121(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=B12;
RA Koskinen R.H., Salomonson J., Tregaskes C.A., Young J.R.,
RA Goodchild M., Bumstead N., Vainio O.;
RT "The chicken CD4 gene has remained conserved in evolution.";
RL Immunogenetics 0:0-0(2002).
DR EMBL; Y12012; CA72740.1; -.
DR EMBL; AJ401223; CAC82027.1; -.
DR InterPro; IPR007110; IG_1-like.
DR Pfam; PF00047; IG_3.
DR PROSITE; PS50835; IG_LIKE; 2.
KW Signal.
FT SIGNAL.
SQ SEQUENCE 487 AA; 54985 MW; B9CBA92EC97F45B CRC64;

```

Query Match 7.0%; Score 240; DB 13; Length 487;
 Best Local Similarity 25.7%; Pred. No. 8.8e-10;
 Matches 113; Conservative 61; Mismatches 184; Indels 82; Gaps 20;

```

Qy      10 LLLVQLALLPAATQGNKVVLGKGDVYELTCTA-SQKSIQPHMK-----SNQIKLG 63
Db      14 VITVLQGLTPVMAQEQEI-GIAGEKVVILSCRAINNQKGTCTWKYKEVSTIISFS 72
Qy      64 NQSGFLTKGPSKLNDRADRSRLMDQNPFLIKNLKLEIDSDPYICEVEDOKEVQLLVF 123
Db      73 KAQVFCKKAPM-----TRSELNSNSKGLKYSDSLDDAGIYTCACYPVVISLHVF 125
Qy      124 GLTANSDTHLLQGSLLTLT-ESPPGSSPS-----VQCRSPRGKNIQG 165
Db      126 KLTISNSGHFLTMEDDELTMQNSHSQPHLSIKLFNINNDIVTTEILQEBAPQ-KYI-- 182
Qy      166 GKTLVSQLELDQSGTYTCTYVLQNKVVEFKI--DIYVLAFOKAS-SIVYKKEEQVBF 222
Db      183 ---LKLQKLAIDSGTWMCHVNSPISNINISFDVAVLGFEXERLEIYITVGTNTAIFS 239

```

```

Qy 223 FPLATVEVLTS--SGELVWQOERASSSSSWITFDLKNKEVSVKRYTQDPKLOMGKPL 28
Db 240 WRLNFRKIKMGKGFPGKLMWEPQGNATIELNFS-----VTHQOELHKTKK-SN 288
Qy 281 HL-----TLFQALP--QVAGSGNLTLAEAKTGKGLHQEVNLVVMRATOL----- 322
Db 289 HIMFEISGKTDGTDGVKIPKQVLNHSGGYKQOLEINGRTSSVRLVVMQYTAIPAGPL 348
Qy 323 ----QKULTCEWGTSPPLMLSTLLEN--KEAKSKREKRVVVLNPPA-GMOC-LIS 373
Db 349 SRGGMTTLTCQVSGPLPSNAHLWBRVNGTQWEMKSKQHEAKVEVNVASPGIMCHIVE 408
Qy 374 DSG-QVLLESNIKVLPTWST 392
Db 409 DNNKISLNTYVEEAHWNS 428

RESULT 89
P70443 PRELIMINARY; PRT; 90 AA.
ID P70443
AC P70443.
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Mutant T-cell surface glycoprotein CD4 (Fragment).
GN CD4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BR/cdJ;
RA Rosenwasser O.A., Fairchild S.P., Tomonari K.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U75219; AAB18133.1; -.
DR HSSP; P05540; ICID.
DR InterPro; IPR007110; Ig-like.
FT NON TER 1
SO SEQUENCE 90 AA; 9938 MW; 6528B707928A4685 CRC64;

Query Match 6.9%; Score 237; DB 11; Length 90;
Best Local Similarity 59.2%; Pred. No. 1.3e-10;
Matches 45; Conservative 14; Mismatches 17; Indels 0; Gaps 0

Qy 310 QEVNLVVMRATOLQKULTCEWGTSPKIMLSLTKENKAKYSKREKRVVVLNPPAGMW 369
Db 1 QEVNLVVMKVAQLNNTLTCEWGTSPKMRITLKQENQARVSEBQKVQVAPETGLWQ 60
Qy 370 CLTSDSGOVLLLESNIK 385
Db 61 CLTSEGDVKVMDSRIQ 76

RESULT 90
Q96RW7 PRELIMINARY; PRT; 5636 AA.
ID Q96RW7
AC Q96RW7.
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hemiscentin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Trent J.;
RA "Human hemiscentin gene.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF156100; AAK68690.1; -.

```

DR GO; GO:0005727; C:extrachromosomal circular DNA; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR000875; Cecropin.
DR InterPro; IPR001434; DUF11.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR005017; GFP_like.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR000169; SHprot_acsite.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR02035; VWF_A.
DR Pfam; PF00008; EGF; 5.
DR Pfam; PF00047; Ig; 44.
DR Pfam; PF00090; tsp_1; 6.
DR SMART; SM00179; EGF_CA; 7.
DR SMART; SM00408; IGC2; 43.
DR SMART; SM00209; TSP1; 6.
DR SMART; SM00327; VMA; 1.
DR TIGRFAMs; TIGR01451; B_ant_repeat; 9.
DR PROSITE; PS00010; ASX_HYDROXYL; 5.
DR PROSITE; PS00268; CECROPIN; 1.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS01187; EGF_CA; 8.
DR PROSITE; PS50835; IG_LIKE; 44.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
DR PROSITE; PS50092; TSP1; 6.
KW EGF-like domain; Immunoglobulin domain.
SQ SEQUENCE 5636 AA; 613660 MW; F000B319CED7B52C CRC64;

```

Query March 5.6%; Score 190; DB 4; Length 5636;
Best Local Similarity 19.5%; Pred. No. 0.00019;
Matches 154; Conservative 107; Mismatch 253; Indels 274; Gaps 38;

QY 34 GDFVELCTTSQKSKSIQPHWKNNSQIKILNGSGFLLTGKSPKLNDRADSRSLMDQGNFP 93
Db 807 GSNVTLCTCYQGYEPPIIKRRLDNMIFSR--PFSVSSISQLTGA----- 851
QY 94 LIINKLIKEDSDTYICEVEDQ---KEEVOLVFGLTA-----NSDTHLLQGSLTTLT 143
Db 852 LFLINLWASDKGYITGAEHQFGKIQGETTYTGVLAFLIGISPSVAANVIEGQQLTLPC 911
QY 144 ESPFGS-----SPSVQCRSPRGKNIOGKTLVSQLELDSDGTWC--- 184
Db 912 TLAGNPIPERRWIKSNAMLLQNYITVRS-----DGLSHERVOLQGGGEYTCVAS 963
QY 185 -----TYLQNKATVEFKIDIVYLAFORAS-----SIYKKEGEVOEF 221
Db 964 NVAGTNNKTTSVVVHVLPTIOHQQLISTIEGIFVTLPCRASGNPKSPVIMSKGELIST 1023
QY 222 SFPLAFLVEKLTGS-GEIIMQOAEBAASSKSMI-----TFDLKNEY-----SV 263
Db 1024 S-----SAKPSAGADGSLIYVSPGEGSGEYVCTATNTAGYAKRKVOLTYVVRPVFQDL 1078
QY 264 KRVTQDEKIQMGKKLPHLTLP---QALP-----OYAGSGNLTIA 300
Db 1079 RGLSQDKPVEISVLAGEVTLPCVKSLLPPPIITMAKETQLISFSPFRHLPBEGSKMIT 1138
QY 301 LEAKT-----GKLHQEVLVVMRATOLQ-----NLTCVEWGP 333
Db 1139 -EFTSDSGMYLCAATNIAQNVVQAVLNLNVHPKIQRGPHLKVQVGQRVDICMANGT 1197
QY 334 TSPKLTM---SLKLENKEAKVSGREKRPVW---LNDEAGMOCLL---SDSGVALL 381
Db 1198 PLPIVITWSKSGSTLVGGEHHVSNPDGTLSDQATPSDAGIYTCVATNIACTDEFTILH 1257
QY 382 -----SNIKVL-PTMSTVEPERKSCDKHTCPRCAP-----ELGG-GPSVFL 421
Db 1258 VQEPFVEDELEPPNYITFOERVANQRIEF-PCPAKGTPTKIKWLHNGRELVTGEPEISI 1316

```

QY 422 FPPPKDILMT--SRTP-----EVTGVVVDV-----HEDPEVK-----453

Db 1317 L---EEGLILVIASTPYDNGEYICVAVNEAGTERKTNLKNHVPVYLKDEQGVNSVYL 1377

QY 454 -----FNMYVDGEVHNAAKTKREEQYNSTYRVSVLTVLHQDWLNG 495

Db 1374 LNOQLTNLFCEVEGTSPPIIMYKDVQVTEST-----IQTVNGKILKLFATPED-----A 1426

QY 496 KEYKQKVSNKA-----LPAPLEKITSKAGQPREQVYTLTPSRHELKNQVS 543

Db 1427 GRYSCKAININGTSOKYFNIDVLVPEPTLIIGT-----NPPKEYSV-----VLRDVA 1477

QY 544 LTCLVKGF-YPSDIAVEMESNGQRPENNYKTPRPVLDSDGSFFLYSKLTVDYKSRMQAGNVF 602

Db 1473 LECQVKGTPRPD---IHMFKDKGRPLFDGDPVVELLDRCQVLHLKNAARRNDKGRYQ----- 1524

QY 603 SCSVIMHEA 610

Db 1525 -CTVSNAA 1531

RESULT	91
Q90524	
ID	PRELIMINARY; PRT; 268 AA.
AC	Q90524;
DT	01-NOV-1996 (TREMBLrel. 01, Created)
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE	Novel antigen receptor (Fragment).
OS	Ginglymosestoma citratum (Nurse shark).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC	Elasmobranchii; Galeomorphi; Galeoidae; Orectolobiformes;
OC	Ginglymosomatidae; Ginglymostoma.
OX	NCBI_TaxID=7801;
RN	[1]
RP	SEQUENCE OF 1-240 FROM N.A.
RC	TISSUE=Spleen;
RX	MEDLINE=95183140; Pubmed=7877689;
RA	Greenberg A.S., Avila D., Hughes M., Hughes A., McKinney E.C.,
RA	Plajnik M.F.;
RT	"A new antigen receptor gene family that undergoes rearrangement and
RT	extensive somatic diversification in sharks.";
RL	Nature 374:168-173(1995).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Spleen;
RA	Greenberg A.S.;
RL	Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
DR	EMBL; U18680; AAB48352.1; .
DR	HSSP; P01842; 7FAB.
DR	GO; GO:0004872; F:receptor activity; IEA.
DR	InterPro; IPR007110; IG-like.
DR	InterPro; IPR003597; IG_c1.
DR	InterPro; IPR003006; IG_MHC.
DR	Pfam; PF00047; Ig; 2.
DR	SMART; SM00407; IgC1; 1.
DR	PROSITE; PS50835; IG_LIKE; 2.
DR	PROSITE; PS00290; IG_MHC; 1.
KW	Receptor.
FT	NON TER
SQ	SEQUENCE 268 AA; 268 MM; OCDEF24DD6E9DFDC4 CRC64;
Query Match	5.4%; Score 186; DB 13; Length 268;
Best Local Similarity	24.3%; Pred. No. 4.9e-06;
Matches	65; Conservative 39; Mismatches 96; Indels 66; Gaps 10;
Oy	417 PSVLFPPKPKDITLMISRTPE-----VTGVVVVDVSHEDPEVKFNMYVDG-----460
Db	14 PYVF-----TARVDDTPQEITKETGESISINCLVRSDSCALPSTYWNKKSGSTNEE65
Oy	461 -----VEVHNNAKTKRREGQYNSTRVAVSLVTLHODMNLNGKEYKKCV--SNKALPAP510
Db	66 TISKGRVETRVNSGSKSFSILIND-----LTIV----EDSGTYRKRYRKWAAYDCG113

```
OY      511 IEXT-----ISXAKGQRRBPYYTL--PPSRBELTKNQVSLTCLVKGFSPSDIA    557  
         :|::||::||::||::||::||::||::||::||::||::||::||::||  
Db       114 LELBLIIYYYGGGTGVTVNPGIPLSPPVSLLSHATTEEQRANGFVOLVCILSGYPENIA   1730  
  
OY      558 VERESNGQP-ENNKYKTPPVLDSDGSFLYSKLITVDKSRMQQGNVFSCSVMEALHNHYT    616  
         ||::||::||::||::||::||::||::||::||::||::||::||:  
Db       174 VSMOKKTKITTSGFATTSSPVTSTNSDFSCASLIRLVLOEWSRGSYSCOVSHATSSNR     233  
  
OY      617 QRSLSLPGLDCTCAADOGDELGLW 644  
        ::||:||::||::||::||::||::||::||::||::||::||::||  
Db      234 KEIKRSTS-----ETAVLRIDPTVEIM 255
```

ID	Q8WU38	PRELIMINARY;	PRT;	573 AA.
AC	Q8WU38;			
DT	01-MAR-2002 (TREMBlrel. 20, Created)			
DT	01-MAR-2002 (TREMBlrel. 20, Last sequence update)			
DT	01-OCT-2003 (TREMBlrel. 25, Last annotation update)			
DE	Hypothetical protein.			
OC	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX	NCBI_TaxId=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Testis;			
RA	Strausberg R.;			
RL	Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; BC021276; AAH21276.1; -			
DR	PIR; S21205; S21205.			
DR	InterPro; IPR007110; IG_1like.			
DR	InterPro; IPR003006; IG_MHC.			
DR	InterPro; IPR003596; IG_V.			
DR	Pfam; PF00047; IG; 4.			
DR	SMART; SMO0406; IGV; 1.			
DR	PROSITE; PS50835; IG_LIKE; 4.			
DR	PROSITE; PS00290; IG_MHC; 2.			
KW	Hypothetical protein.			
SQ	SEQUENCE 573 AA; 62967 MW; FD07234403JACS30 CRC64;			
Query Match	5.4%;	Score 184.5;	DB 4;	Length 573;
Best Local Similarity	21.0%;	Pred. No. 1.9e-05;		
Matches 151;	Conservative 77;	Mismatches 248;	Indels 243;	Gaps 34
QY	1 MNRGVPPFHLILV---QLALLPAATQGNKVLVGGKGTVELETCTAS--QKKSIGFHWKN 55			
DB	1 MELGLSWFLFLAILIKGVQCEVQLVESGGGLV---QPGKSLILSCAASFTPDYDMHWVR 57			
QY	56 SNOIK-----ILGNQSF-----LTKGPKSLNDRASRSRLMDQGNFLIIKNLKIEDS 104			
DB	58 QAPKGLIEWWSGISMNSGIYADSVKGRFTIS--RDNAKNSLYIQMN-----SLRAEDT 110			
QY	105 DTYICEVBDQKEVQLVFLGLTANSDTHLLOGSITLTLESPPGSSPSV-----QCRSPR 159			
DB	111 ALYYCAKHGSGSYI-----GYYGMD--VWGCGTTLVVSAPTKAPDVFPPIISGCRHPK 162			
QY	160 GKN-----IQGGKTLVSQLELDQSGTWTCTVLQNGKKEVFKIDIVLAFQKASSTLY 212			
DB	163 DNSPVLACLITGTHPTSTV---TW-----Y 186			
QY	213 KKEGEQVEYFSP-----LAFVTEKLTGSGELMWQAE-----RASSKSWTFDLKNKE 260			
DB	187 MGTGSGQGRTPPELQGRDSYMTSSQLSTPLQMRQSEYKCVNGHTAKS-----KKE 239			
QY	261 VSVKRVTPDPLQMGKKLPLHLTLPLALPOYAGS---GNLTALAEAKTGKTLQEVNLVVM 317			
DB	240 IF--RWPESPFAQAS-----SVTPAQPOAQEGSLAKATTA PATTRNTGRGSEE----- 284			
QY	318 RATLQKNLICEVWGPTSPKMLMSLKLENKAKYSKREKPVWVLNPEAGMOCLLSDSQ 377			


```

Db 285 -----KKKEKEKEOEERE----- 298
Qy 378 VLESNIKVLPTMSTFPERPKSCDKHTTCTPCCAPPELLGSPVFLPPPKKOTLMISRTPE 437
Db 299 -----TKTPECSTHTQP--LG---VYLLTPAVQD-LMLRDKAT 330
Qy 438 VTCVVVDVSHEDPEVKFNMYVDG-VEVNAKTKPREEQNTSTYRVSVLTVJHQPWLNGK 496
Db 331 FTCFVVGSLDKDAHL--TWEVAKVPTGVEGGLERHNSGQSCHSRLLTLRSLMNACT 388
Qy 497 EYKCKVSNKALP-----APIKRTIS-KAKGQPREPPQVYTLTPPSDELTQKQV 542
Db 389 SVTCTLMHPSLPPORLMALREPAQAQPVXLISLILASSDPPEASV----- 434
Qy 543 SLTCLVKGFPYPSDIAVEMESNGQPENN---YKTTTPVLDSDSGFLYSLKLTVDKSRMOG 599
Db 435 -LLCEVSGFSPPNILMLLEDEQREVNTSGFAPARPPQSGTTFAMVSLVLRPAPSPPP 493
Qy 600 NVFSGSVHME---ALHNHYTQKSLSLSPGLQDETCAEAOQDELQ-----GLMTT 646
Db 494 ATYTCTVSHEDRTLLN--ASRSLEVS-YLMTPLPLPSKDNSDDYTFDDVGLSWTT 549

```

RESULT 93

Q9DBM4

PRELIMINARY; PRT; 130 AA.

```

ID 09DBM4
AC 09DBM4
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE 1810027001R1K
GN 1810027001R1K
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Pancreas;
RX MEDLINE=21085660; PubMed=11217811;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamamata I.,
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kankawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Giesl C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nishida I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.T., Bult C., Fletcher C., Fujita K., Gariboldi M.,
RA Guelincich S., Hill D., Holtman M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Norioka P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wymah-Botie A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK007622; BAB25142.1; -.
DR PIR; S22760; S22760.
DR HSSP; P01842; 7FAB.
DR MGD; MGI:1916426; 1810027001R1K.
DR InterPro; IPR007110; IG-1like.
DR InterPro; IPR003597; IG_C1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00407; IgC1.1.
DR PROSITE; PS00835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
SQ SEQUENCE 130 AA; 14253 MW; 438197975E76E54 CRC64;

```

```

Query Match 5.4%; Score 183.5; DB 11; Length 130;
Best Local Similarity 36.5%; Pred. No. 2.7e-06;
Matches 38; Conservative 18; Mismatches 43; Indels 5; Gaps 3;
Qy 520 GQPR-EPQVYTLPPSRDELTKQVSLTCLVKGFPYPSDIAVEMESNGQP-ENNYKTTTPVL 577
Db 25 GQPKSPSVTLFPSPSELETNKATLVCTITDFYGVGVTVDMKVDTGVTQGMETTPQSK 84
Qy 578 DSDSGFLYSLKLTVDKSRMOQGNVPSGVMEDALHNHYTQKSLS 621
Db 85 QSNMKYMASSYLTLTARAMERHSSYSQVTHE--GHTVEKSL 125

```

RESULT 94

Q8N355

PRELIMINARY; PRT; 234 AA.

```

ID 08N355
AC 08N355
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (Apr-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC028090; AAH28090.1; -.
DR PIR; S12441; S12441.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-1like.
DR InterPro; IPR003597; IG_C1.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; Ig_2.
DR SMART; SM00409; IG; 2.
DR SMART; SM00407; IG; 1.
DR SMART; SM00406; IG; 1.
DR PROSITE; PS00835; IG_LIKE; 2.
DR PROSITE; PS00290; IG_MHC; 1.
KW Hypothetical protein.
SQ SEQUENCE 234 AA; 24792 MW; CC848CAEBA49D63 CRC64;

```

Query Match

5.4%; Score 183; DB 4; Length 234;

Best Local Similarity 23.9%; Pred. No. 6.8e-06;

Matches 50; Conservative 42; Mismatches 99; Indels 18; Gaps 6;

```

Qy 415 GQPSVFLFPKPCKOTLMISRTPEVTCVVVDVSHEDPEVKFNWY-----VDGVEYNAKT 468
Db 16 GSVSYVLTQPPSVAVAGQTAITCCGNNIGSK-----SVHMYQKPGQAFLVLYDDSD 71
Qy 469 KPR---EEQNVSTYRVSVLTVLVHQDMLNGEKYCKVSNKALPAPI---EXTISKAKGQ 522
Db 72 RPSGIPERFSSNGSGNTATLTISRVDAGDEADYCCQLDSSDHPVFGGTFKLTVLQCP 131
Qy 523 R-EPQVYTLPPSRDELTKQVSLTCLVKGFPYPSDIAVEMESNGQP-ENNYKTTTPVLSD 580
Db 132 KAAPSVTLFPSPSELETNKATLVCTITDFYGVGVTVDMKVDTGVTQGMETTPQSK 191
Qy 581 GSPFLYSLKLTVDKSRMOQGNVPSGVMHE 609
Db 192 NKYAASVSLTPQMKSHRSYSCQVTHE 220

```

RESULT 95

Q28260

PRELIMINARY; PRT; 739 AA.

```

ID 028260
AC 028260
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

```

DT 01-OCT-2003 (TREMBLERel. 25, last annotation update)
DE Vascular cell adhesion molecule-1.
OS *Canis familiaris* (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Carnivora; Fissipedia; Canidae; *Canis*.
OX NCBI_TaxId=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Ballantyne C.M., Clubb F.J., Perrard J.L., Radencovic B.,
RA Youker K.A., Smith C.W., Entman M.L., Hawkins H.K., Frazier O.H.,
RA Willerson J.T.;
RT "Increased Expression of VCAM-1 and ICAM-1 in Early Cardiac Allograft
RT Arteriosclerosis in the Dog."
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U32086; AA84866.1; -.
DR HSSP; P19320; 1VCA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0016337; P:cell-cell adhesion; IEA.
DR InterPro; IPR003987; ICAM VCAM-1.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_G2.
DR InterPro; IPR003989; VCAM-1.
DR Pfam; PF00047; Ig_7.
DR PRINTS; PR01472; ICAMVCAM1.
DR PRINTS; PR01474; VCAM1.
DR SMART; SM00408; IGC2_4.
DR PROSITE; PS50835; IG_LIKE; 5.
KW Immunoglobulin domain.
SQ SEQUENCE 739 AA; 81412 MW; BB5DA3853739C615 CRC64;

Query Match	5.3%;	Score 182.5;	DB 6;	Length 739;
Best Local Similarity	21.5%;	Pred. No. 3.8e-05;		
Matches 112;	Conservative 96;	Mismatches 183;	Indels 131;	Gaps 29;

```

Db 237 QEGSGVTMTCSSEGLPPFOIFMSKLDNGNLOLL----- 270
Oy 89 QGNPFLIIKNKLTIEDSDTYIC---EVEDQKEEVOLLV---FGLTANSDTHLO-QGS 138
Db 271 SGNATLTLIARLEDSGTVCGEVNEVGKDGEVLLIYOEKFPFVIEISPGOIIAQIDS 330
Oy 139 LTLTLESPPGSSPS-----VOCSPRG--KNIOGK-TLSVSQLEODSGTMTCTYLOQK 191
Db 331 VLTGCVTDCSPSPSMTQIDSLPSGTVCYKKGAKSTILTSVNLNENHSYLCYTCGHK 390
Oy 192 KYEFKIDIVLAFOK-----ASSIYKKEGEVPEPSFLAFTVEKLTSGGELMOAERAS 246
Db 391 KLEKGIKVDLVSFPDPREVMESGLLVGDNPLTVSCEVENVPDSRL-EIEL-FKGETII 447
Oy 247 SSKSMI-----TFLPNKKEVSVKRVNTOQPKLOMKKL---PLHLTLPLQALPOVAGSN 296
Db 448 ESKSFLIEDMDKKSLTSLKLEMTFIPITBD---TAKVLVCLAKLHIDEMEFPKORGS- 501
Oy 297 LTLAEATGKXLHQEVNLYVMRATOLQK---NLCEVMGPTSPKMLSLKLENAEVVS 352
Db 502 -TQTLVYVAV-RDTTVVSPSSIVEGSPYNNMTCSSDGLPAPNIIIMSRRLSNRLO-S 557
Oy 353 KREKRVWVLN---PEAGMOQL-LSDSQVLLSENIKVLPLWSTFVPEKSDCKTHTCPP 407
Db 558 LSEDDILTLTSAKMEDSGIYCEGINQAG-TSRKEVELI---IQVAPKDI----- 603
Oy 408 CPAPELLGSPSYFLPPPKPOTLMSIRPEVYCVVADVSHEDPEYKFMWYV--DGEVYHN 465
Db 604 ----QLINFPSSV--KEGDVIIS---CTC-----GNVPKMTIILKKAELFGD 643
Oy 466 AKTKPREQYNSTYRVVSVLTVLHODML-NGKEYICKVSNKA 506
Db 644 TVLKSRLDAY-----TIHVOLEDAGVYCESKNEA 674

```

RESULT 96
Q63669

ID	063669	PRELIMINARY;	PRT; 739 AA.
AC	063669;		
DT	01-NOV-1996 (TREMBLrel_01, Created)		
DT	01-NOV-1996 (TREMBLrel_01, Last sequence update)		
DT	01-OCT-2003 (TREMBLrel_25, Last annotation update)		
DE	Vascular cell adhesion molecule 1 precursor.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;		
OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
RN	NCB1_Taxid=10116;		
RP	[1]		
RC	SEQUENCE FROM N.A.		
RC	STRAIN-Sprague Dawley; TISSUE-Lung;		
RX	MEDLINE=92305064; PubMed=1377031;		
RA	Williams A.J., Atkins R.C., Fries J.W., Gimbrone M.A.Jr.,		
RA	Cybulski M.I., Collins T.;		
RL	"Nucleotide sequence of rat vascular cell adhesion molecule-1 CDNA.";		
RT	Biochim. Biophys. Acta 1131:214-216(1992).		
DR	EMBL; X63722; CAA45254.1; -.		
DR	HSSP; P19320; IVCA.		
DR	GO; GO:0016020; C:membrane; IEA.		
DR	GO; GO:0016337; P:cell-cell adhesion; IEA.		
DR	InterPro; IPR003987; ICAM VCAM-1.		
DR	InterPro; IPR007110; IG-like.		
DR	InterPro; IPR003599; IG c2.		
DR	InterPro; IPR003989; VCAM-1.		
DR	pfam; PF00047; Ig; 5.		
DR	PRINTS; PRO1472; ICAMVCAM1.		
DR	PRINTS; PRO1474; VCAM1.		
DR	SMART; SMO0408; IGc2; 4.		
DR	PROSITE; PS50835; IG LIKE; 5.		
KW	Immunoglobulin domain; Signal		
FT	SIGNAL 1 24		POTENTIAL.
FT	CHAIN 25 739		VASCULAR CELL ADHESION MOLECULE 1.
SEQ	SEQUENCE 739 AA; 81136 MW; ASAAD112F67FB96 CRC64;		

Query Match	5.3%;	Score 182.5;	DB 11;	Length 739;
Best Local Similarity	20.8%;	Pred. No. 3.8e-05;		
Matches 124;	Conservative 87;	Mismatches 195;	Indels 191;	Gaps 28;

```

QY      33 KGDYVLTCTASQKKKSIOFHMKNSNOIKILGNQSGFLTKGSSKLNDBADSRSLMDQGNF 92
Db      238 EGAATMTKASBGLPAPFIWMSK-----LDNGVQLL-----SGNA 274
QY      93 PLIKNKLIEDSDPYICE---VEDOKBEVOLV-----FGLTANSDPHLLOQSILTLT 142
Db      275 TLTLIARMEBDSGIYCEGVNLVGRDKTEVELIYQEKPFYVIDBSGSAVAQVDSVLT 334
QY      143 LESPPGSSPSYQCCSPRKNIOG-----GKITSVGOLELODSGTWTCTYLQNOKYEF 195
Db      335 CAAVGCDSPSSMRWRTQDTSPLNGVRDEGATSTLTSPVGEDBSHSLCYTTCORRLEK 394
QY      196 KIDIVLAFQASSIIVYKKEGOVEFSPPLAFVTEKLTGSGELMMQABRASSKSMITFD 255
Db      335 TIQEVYSF-----PEDPELIEISPLVH----- 417
QY      256 LKNKEVSYKRVTDPKLQMGKKLPLHLTLPOALFOYAGSGLTLALBAKTKLHOEVNLY 315
Db      418 -----GRPVYNTCTPVNVYFED-----HLEIELL 441
QY      316 VMRATOLOKULTCSEVGPSTPKMLSLKLENKEAKVSKREKPVVNLNPEAGMMOCLLSDS 375
Db      442 KGETTLNKKFLPEEL-GTKS-----LEKYSLEMT-----FTPTA-----EDT 477
QY      376 GOVLLESNIKULPTMSPTFVPEBKSCDKHTTCCPPCAPELLGSPSVFLFP-PRKXDTLMISR 434
Db      478 GKALV--CLATLHSSQMESEPKQKQSOYTLVYVNAFK--EPTLVMSPSVPBEGSPVYN 531
QY      435 TPBYTGVVVDVSHDEPVEKENVYDQVEVHNAKTKPREBOYNSYTRVVSULTVLHODMLN 494
Db      532 ---LTC-----SSDGFPPPKILM---SRQLKNGELQLOPSQ---NTT---LSFNATKXED--- 573

```

Db 532 ---LTC---SSDGFPTPKILM---SRQKNGEIQLPSQ---NTT---LSFMATKMD--- 573

Qy	495	GKEYCKXSNKALPAIEKTIK	-----AKQPEPOVYTLPPSDELTKNOVSLTCLV	548	
Db	574	SGIYCEGINEA	-----GISKSVELLIOGSSKOTIQ-TVPSSKVSKEGTIVIIISCTC	625	
Qy	549	KGFYPSDIAVENESNGOPENNYKTPPVLD	S-DGSFFLTKSLTVDSRMOQGNVESC	604	
Db	646	-GSVPEIWI	-----LKKKAKTGDWLVKSYNGSI	-----TIRKALQIDAGVYEC	668
RESULT 97					
ID	Q95YM2	PRELIMINARY:	PRT: 17352 AA.		
AC	Q95YM2				
DT	01-DEC-2001	(TREMBLrel. 19, Created)			
DT	01-DEC-2001	(TREMBLrel. 19, Last sequence update)			
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)			
DE	I-connectin.				
GN	I-CON.				
OS	Procambarus clarkii (Red swamp crayfish).				
OC	Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;				
OC	Eunalaecostiracae; Eucarida; Decapoda; Pleocyemata; Astacidea;				
OC	Astacoidae; Cambaridae; Procambarus.				
OX	NCBI_TaxID=6728;				
RL	EMBO J. 20:4826-4835(2001).				
CC	-1 SIMILARITY: CONTAINS 1 SH3 DOMAIN.				
RA	MEDLINE=21423462; PubMed=11532946;				
RA	Fukuzawa A., Shimamura J., Takemori S., Kanzawa N., Yamaguchi M.,				
RA	Sun P., Maruyama K., Kimura S.;				
RT	"Invertebrate connectin spans as much as 3.5 micrometer in the giant				
RT	sarcomeres of crayfish claw muscle."				
DR	InterPro: IPR003957; FGGY kin.				
DR	InterPro: IPR008957; FN_III-like.				
DR	InterPro: IPR007110; IG-1-like.				
DR	InterPro: IPR003598; IG_c2.				
DR	InterPro: IPR01452; SH3.				
DR	Pfam: PF00041; fn3; 5.				
DR	Pfam: PF00047; Ig; 49.				
DR	Pfam: PF00018; SH3; 1.				
DR	SMART; SM00408; IGC2; 13.				
DR	SMART; SM00326; SH3; 1.				
DR	PROSITE; PS00933; FGGY_KINASES_1; 3.				
DR	PROSITE; PS50835; IG_LIKE; 49.				
DR	PROSITE; PS50002; SH3; 1.				
DR	Immunoglobulin domain; SH3 domain.				
SO	SEQUENCE 17352 AA; 1962348 MW; 4BA157BEC042E42D CRC64;				
Query Match 5.3%; Score 182; DB 5; Length 17352;					
Best Local Similarity 19.9%; Pred. No. 0.0038;					
Matches 128; Conservative 103; Mismatches 227; Indels 184; Gaps 29					
Qy	37	VELTCTASOKKSIOHFMKNSNOIKILINGOSFELTKGPSKLANDRADSRSLMDQGNFLII	96		
Db	16247	VEVT---NEKAVTVMH-KDGEITEHHEKFKLVSGGKTR-----KLVL	1622		
Qy	97	KNLKIEDSDTYICEVDDQKEVQLVFGILTAN-----SPTLLQGOSLTLLSPSPGSS-	150		
Db	16286	MEATLSDSGEYTCVLDQECTMETLVRELPAEIVKMKQOVVSKGRATMEVELTGTDAV	1633		
Qy	151	-----PSVOCR-----SPRKNIGGKTLVSQLELDSGMTCVTVLQNKVEFKID	198		
Db	16346	ITWYDVEVIRSDHYKSLSDGK---VORLVNVCQFDSGTYRAVVKSECSATLVKE	1644		
Qy	199	IVVLAFOKASSIVYKKEG-----QVESFPLAFTVEKLTGSGELMW---QAKRAS	246		
Db	16402	LQV-----EGDFSKKLPAQMDVNFKTATTFVE-ITKDYEVMKMLREAGELSS	1644		
Qy	247	SSKSVITTDLKKKEVSVKRVTDPPKIQ-----NGKKLPLHLTLPOALPOYAGSGNTLT	299		

Db	16448	SEKVIYKGEAKKRKILIVKVSQCDNAEYSVCLGNLKTSCVLHVAMETAP-----	16497
Qy	300	ALEAKTGLKHQEVNLVVMRATOLQKNLTCV--WGPTSKLM-----ISLKLENK---A	349
Db	16498	----KIPEKHQKEVIAVITKG-----KDAVLKVPFTATPTPKVPMYHKGQLINTENTKLLP	16549
Qy	350	KVSKREKRPV---WYLNPEAGMOC--LLSSGOVLLESNIKVLPMTSTPVEPSSCKHTTC	405
Db	16550	TISQKESITIKQVENIDGCEYKLLCNQGAAYADFTLILKDKSQPQETPEMVEVTN-	16608
Qy	406	PCPCAPELLGSPVFLFPFKPKOTLMSRTPETCVVVDVSHEDVEKMWVVDGVEVHN	465
Db	16609	-----SVTLHMSLPK-----DGRALTIIVYMSKKQ---SWVFNQGV--	16647
Qy	466	AKTKPREEOYNSTYRVSVVLVLADQMLNGKEYCKKVSNMKALPAPIEKTISKAGQPREP	525
Db	16648	-----KITQSTIVVTSILTT--HE-----EYCFPRVS-----AENIGRSEASNASR	16685
Qy	526	QVYTLPPSSBELTKNOVSLTCLVKGYPBDI-----AVEMESNGQEN-----	568
Db	16686	YKVCCEVRAEPVPVKEOLOAVTGHQHEVIRACVTATPTPKIPLMKDKGEVSGNTSYE	16745
Qy	569	NYKTTPEVLDDSGSFFLYSKLLTVDSKRMQGNVFSQSWHEA	610
Db	16746	NFRAT-----LTIKETMTSGGMYTCRASNEA	16772
RESULT 98			
08ISF3	08ISF3	PRELIMINARY; PRT; 2693 AA.	
AC	Q8ISF3		
DT	01-MAR-2003	(TREMBLrel. 23, Created)	
DT	01-MAR-2003	(TREMBLrel. 23, Last sequence update)	
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)	
DE	301kDa_2	protein.	
GN	ISOP.		
CS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditia; Rhabditoidea;		
OC	Rhabditidae; Pelodierinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=22269627; PubMed=12381307;		
RA	Fidelity D., Gernert K., Shmeleva N., Tang X., Mercer K.,		
RA	Borodovsky M., Benan G.;		
RT	"Titins in Caenorhabditis elegans with Unusual Features: Collod-coil		
RT	Domains, Novel Regulation of Kinase Activity and Two New Possible		
RT	Elastic Regions.";		
RL	J. Mol. Biol. 323:533-549(2002).		
DR	EMBL; AY130758; AAN61521.1; "		
DR	GO; GO:0005524; F:ATP binding; IEA.		
DR	GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.		
DR	GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.		
DR	GO; GO:0006468; P:protein amino acid phosphorylation; IEA.		
DR	InterPro; IPR003961; FN_III.		
DR	InterPro; IPR007110; IG_1like.		
DR	InterPro; IPR003598; IG_C2.		
DR	InterPro; IPR000719; Prot_kinase.		
DR	InterPro; IPR002290; Ser_Thr_kinase.		
DR	InterPro; IPR008271; Ser_Thr_kin_AS.		
DR	InterPro; IPR001245; Tyr_kinase.		
DR	Pfam; PF00041; fn3; 2.		
DR	Pfam; PF00047; ig; 19.		
DR	Pfam; PF00069; pkinase; 1.		
DR	ProDom; PD000001; Prot_kinase; 1.		
DR	SMART; SM00060; FN3; 1.		
DR	SMART; SM00408; IGC2; 19.		
DR	SMART; SM00220; S_TKc; 1.		
DR	SMART; SM00219; TyKc; 1.		
DR	PROSITE; PSS00835; IG_LIKE; 18.		
DR	PROSITE; PSS00107; PROTEIN_KINASE_ATP; 1.		
DR	PROSITE; PSS00011; PROTEIN_KINASE_DOM; 1.		
DR	PROSITE; PSS0108; PROTEIN_KINASE_ST; 1.		

```

SQ SEQUENCE 2693 AA; 299094 MW; 1CE4AD2011E8391A CRC64;
Query Match 5.3%; Score 180.5; DB 5; Length 2693;
Best Local Similarity 20.2%; Pred. No. 0.00034;
Matches 137; Conservative 99; Mismatches 254; Indels 189; Gaps 33;

QY 28 VLKGGKDTVELTCTASQKSIQPHWKNXNQIKILGNQSFLLTKGPSKLNDRADSRSLW 87
DB 1100 VLKTAGETATFTCGSYANPAQVWV-----LHNGKALQQTNSYKTRLF 1144
QY 88 DQGNFLIKKLKIEDSTYICEVEDQKEV-----QLLVGLTANS----- 129
DB 1145 DDMATLIVENVDELCTGTAVANNQGDVHTSAQLTISGEAKKIASLPYFIIEIKP 1204
QY 130 DTHLLQGSLLTLESPPGSSP-----SVQCRSPR-----GKNIQSGKTLVSQ 173
DB 1205 KINVEG--ATLSIQADLNGSPPIREVWLNKNSLVEDRLOMKCDGNYQ-----LVVD 1258
QY 174 LELQDSGTWCTVLAQNKVEFKIDIVLAFQKASSIVYKKEGQVEF-----S 222
DB 1259 VGLDEGTYYITAEENKGIKQNTVEVSV---TKSEYKKEKKEKKVEKDEGKKKPGPG 1315
QY 223 FPL---AFTVEKLT-----GSGELMWQAEPRASSKSWTF-DLNKEVSVK----- 264
DB 1316 LPRPSGASKTEQVTFAPDAPSEGPADSYEVERRCPDQREWVSCGSTSLLEIKGLTPNT 1375
QY 265 ----RVTDPKLQMGKKLPLHLTLPOA----LPQYAGSGNLTALAEAKTKLHOENVLV 316
DB 1376 EYIRVAGKNQKGLGEMSEMTSLTKTASVGAPOF-----TISPSK-----IIA 1420
QY 317 MRATQLOKNTLCEVWGTPSPKMLSLKLENKAKVSKREK-----PVWLVN---PEA 365
DB 1421 NRDEFE--IAVEFSGTPTP-----SVKWKENLQVPEKIDVATTSTSSILNKSQEN 1474
QY 366 GMMQCLL-SDSGOVLLESNIKVLPTWSTPVEPKSCDKT--HTCPPCAPPL----- 413
DB 1475 GTFNCLEIENLGAQASASCOVTIFN-----KPSLQSTPHSLERNLVPFLQKALNESA 1528
QY 414 LGSPSVFLFPKPKDITLMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVNAKTKPRE 473
DB 1529 QAGQIWL-----TCRISRSSESTVA-----WPKDDRISAGRYELSS 1567
QY 474 QYNSTYRVSVLVTLHODMLNGEKYCKVSNKALPAPIEKTISAKAQPREPQVYTLPPS 533
DB 1568 DKSNHKLIV-CHAVQSD--TGR-YRCVVTNKYGAESECNVAV-----EDVTKFIAPS 1617
QY 534 -----RDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLDSDGSFELY 586
DB 1618 PSATISDSTALIGHNITLCEKVEGSPAPE--VSWTKQGERLSTTRIRLQTDGENG--- 1671
QY 587 SKLITVDSKRWQGNVFS 605
DB 1672 CKLSISKAESDMDGVVCS 1690

RESULT 99
Q8ISF4 PRELIMINARY; PRT; 2708 AA.
Q8ISF4 AC O8ISF4;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE 301kDa_1 protein.
GN ISOF.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22269627; PubMed=12381307;
RA Flaherty D., Gernert K., Simeleva N., Tang X., Mercer K.,
RA Borodovsky M., Benian G.;

```

```

RT "tilins in Caenorhabditis elegans with Unusual Features: Coiled-coil
RT Domains, Novel Regulation of Kinase Activity and Two New Possible
RT Elastic Regions."
RL J. Mol. Biol. 323:533-549(2002).
DR EMBL: AY130758; AAN61520.1; -.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO: GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR GO: GO:0004648; P:protein amino acid phosphorylation; IEA.
DR InterPro: IPR003961; FN III.
DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR003598; Ig_c2.
DR InterPro: IPR000719; Prot_kinase.
DR InterPro: IPR002290; Ser_thr_kinase.
DR InterPro: IPR008271; Ser_thr_kin_AS.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam: PF00041; fn3_2.
DR Pfam: PF00047; Ig_19.
DR Pfam: PF00069; pkinase; 1.
DR ProDom: PD000001; Prot_kinase; 1.
DR SMART: SM00060; FN3; 1.
DR SMART: SM00408; IG2; 19.
DR SMART: SM00220; S_TKC; 1.
DR SMART: SM00219; TYRC; 1.
DR PROSITE: PS50835; IG_LIKE; 18.
DR PROSITE: PS50107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
SQ SEQUENCE 2708 AA; 300729 MW; 2BEBE8656FA571 CRC64;

Query Match 5.3%; Score 180.5; DB 5; Length 2708;
Best Local Similarity 20.2%; Pred. No. 0.00035;
Matches 137; Conservative 99; Mismatches 254; Indels 189; Gaps 33;

QY 28 VLKGGKDTVELTCTASQKSIQPHWKNXNQIKILGNQSFLLTKGPSKLNDRADSRSLW 87
DB 1115 VLKTAGETATFTCGSYANPAQVWV-----LHNGKALQQTNSYKTRLF 1159
QY 88 DQGNFLIKKLKIEDSTYICEVEDQKEV-----QLLVGLTANS----- 129
DB 1160 DDMATLIVENVDELCTGTAVANNQGDVHTSAQLTISGEAKKIASLPYFIIEIKP 1219
QY 130 DTHLLQGSLLTLESPPGSSP-----SVQCRSPR-----GKNIQSGKTLVSQ 173
DB 1220 KINVEG--ATLSIQADLNGSPPIREVWLNKNSLVEDRLOMKCDGNYQ-----LVVD 1273
QY 174 LELQDSGTWCTVLAQNKVEFKIDIVLAFQKASSIVYKKEGQVEF-----S 222
DB 1274 VGLDEGTYYITAEENKGIKQNTVEVSV---TKSEYKKEKKEKKVEKDEGKKKPGPG 1330
QY 223 FPL---AFTVEKLT-----GSGELMWQAEPRASSKSWTF-DLNKEVSVK----- 264
DB 1331 LPRPSGASKTEQVTFAPDAPSEGPADSYEVERRCPDQREWVSCGSTSLLEIKGLTPNT 1390
QY 265 ----RVTDPKLQMGKKLPLHLTLPOA----LPQYAGSGNLTALAEAKTKLHOENVLV 316
DB 1391 EYIRVAGKNQKGLGEMSEMTSLTKTASVGAPOF-----TISPSK-----IIA 1435
QY 317 MRATQLOKNTLCEVWGTPSPKMLSLKLENKAKVSKREK-----PVWLVN---PEA 365
DB 1436 NRDEFE--IAVEFSGTPTP-----SVKWKENLQVPEKIDVATTSTSSILNKSQEN 1489
QY 414 LGSPSVFLFPKPKDITLMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVNAKTKPRE 473
DB 1490 GTFNCLEIENLGAQASASCOVTIFN-----KPSLQSTPHSLERNLVPFLQKALNESA 1543
QY 414 LGSPSVFLFPKPKDITLMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVNAKTKPRE 473
DB 1544 QAGQIWL-----TCRISRSSESTVA-----WPKDDRISAGRYELSS 1582
QY 474 QYNSTYRVSVLVTLHODMLNGEKYCKVSNKALPAPIEKTISAKAQPREPQVYTLPPS 533
DB 1583 DKSNHKLIV-CHAVQSD--TGR-YRCVVTNKYGAESECNVAV-----EDVTKFIAPS 1632

```

QY 534 -----RDELTKNQVSLTCLVKGFYPSDIAVEMESNGQPENNYKTTPPVLDSDGSFPLY 586
DB 1633 FSATLSDSTAILGHNITLLECKVGEPSAPE--VSWTKDGERISTTRIRIQTQDENGN----- 1686
QY 587 SKLTVDKSRMQGNVFSQS 605
DB 1687 CKLSISKAESDDMGVYVCS 1705
RESULT 100
Q8MNS0 PRELIMINARY; PRT; 2780 AA.
AC Q8MNS0;
DT 01-OCT-2002 (TREMblrel. 22, Created)
DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Hypothetical protein.
GN F12F3.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Br1stol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Waterston R.;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Br1stol N2;
RA Fulton B., Wohlmann P.;
RT "The sequence of C. elegans cosmid F12F3.";
RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Br1stol N2;
RA Waterston R.;
RL Submitted (MAY-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; U80022; AAA29672.1; -.
DR WormRep; F12F3.2a; CE30753.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0004674; P:protein serine/threonine kinase activity; IEA.
DR GO; GO:0016740; P:protein-tyrosine kinase activity; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR003962; FNIII subd.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR007110; IG-1-like.
DR InterPro; IPR003598; IG_C2.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR008271; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00047; ig; 19.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00014; FNTYPEIII.
DR PRODOM; PD000001; Prot_kinase; 1.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00408; IG2; 19.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS50835; IG_LIKE; 18.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW Hypothetical protein; ATP-binding; Immunoglobulin domain; Repeat;
KW Transferase.

SEQ SEQUENCE 2780 AA; 308646 MW; 046D057107B935C1 CRC64;
Query Match 5.3%; Score 180.5; DB 5; Length 2780;
Best Local Similarity 20.2%; Pred. No. 0.00036;
Matches 137; Conservative 99; Mismatches 254; Indels 189; Gaps 33;
QY 28 VLGGKGPVVELTCTASQKSIQFMKNSNQIKILGQGSFLTXGPKLNDRADRSRLW 87
DB 1187 VLKTAGEETATFTQSYANPAQYVW-----LHGKALQGTQSKYKTRLF 1231
QY 88 DQGNPLIIKNIKIEDSDTYICEVEDQEEV---QLLVGLTNS----- 129
DB 1232 DQNTATLVIEENVTELCGTAYAVANQFGDVHTSAQLTSGSEAKKIAASLPYFIELKP 1291
QY 130 DTHLLQGSLLTLTLESPPGSSP-----SYQCSPP-----GKNIQSGTLTYSQ 173
DB 1292 KINVEG--ATLSTIQADLNGSPIPEVWLKUNSLVESDRIQMKCDGVNYO---LIVRD 1345
QY 174 LELQDSGTWCTCTVQNGKVEFKIDIVLAFQKASSIYKKGEQVEF-----S 222
DB 1346 VGLDEGTYITTAENEKGIQNTENVSV---TKSEYKEKKEKKVEKDGKKKPKRPG 1402
QY 223 FPL--AFTVEKLT-----GSGELMWQAEKASSSSKSWITF-DLKNKEVSK----- 264
DB 1403 LPRPSGASKTEQVTMAFPAPSEGPADSVEYERRCPDQREWVSCSTYSLEILEIKLTPNT 1462
QY 265 ----RYQDPRLQMGKPLHLTLPPA---LPQYASGNLTLLAEKTKGLHQEVNLV 316
DB 1463 EYIFRVAGKNQIGGESEMTSTIKTASVGAQPF-----TISPSK-----TIA 1507
QY 317 MRATOLQNLTCFVWGPTSPPLMLSLKLENKAYSKREK-----PWLNL---PEA 365
DB 1508 NRDEFE--IAVPSGTPF-----SVKRYKENLQIVPEKIDVATTSSTLNLKSGDEN 1561
QY 366 GMMOCLL-SDSGQVLLSNNIKVLPWTSTPYEPKSCDKT--HTCPPAPAEI----- 413
DB 1562 GTFNCLIEINELGQASASCOVTFIN-----KPSLQSTPDHSLBRNLVPTLQKALNNEA 1615
QY 414 LGGSVPLFPKPKDITLMIISTPEVTCVVVDVSHEDPEVKKNVYVDGVEVNAATKPRE 473
DB 1616 QAGQIML-----TCRISRSSESTVA-----WRDDEIRISAGRYELSS 1654
QY 474 QYNSTYRVSVLTVLHODMLNGKEYKCKVSKALPAPIEKTISKAKGPRPQVYTLPPS 533
DB 1655 DKSNHKLV-CHAVSQD--TGR-YRCVYTKKGYAESECVAV-----EDYKFIAP 1704
QY 534 -----RDELTKNQVSLTCLVKGFYPSDIAVEMESNGQPENNYKTTPPVLDSDGSFPLY 586
DB 1705 FSATLSDSTAILGHNITLLECKVGEPSAPE--VSWTKDGERISTTRIRIQTQDENGN----- 1758
QY 587 SKLTVDKSRMQGNVFSQS 605
DB 1759 CKLSISKAESDDMGVYVCS 1777

Search completed: August 3, 2004, 13:13:21
Job time : 55.4955 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 12:50:09 ; Search time 65.926 Seconds
(without alignments)
2777.216 Million cell updates/sec

Title: SEQ3
Perfect score: 3414
Sequence: 1 NMRGVFRLHLVLQLALP.....DETCALAQDGLGWTDP 648

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 125 summaries

Database : A_Geneseq_29Jan04:*
1: geneseq1980a:*
2: geneseq1990a:*
3: geneseq2000a:*
4: geneseq2001a:*
5: geneseq2002a:*
6: geneseq2003a:*
7: geneseq2003be:*
8: geneseq2004a:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3268.5	95.7	631	1 AAP93009	Aap93009 Genetic c
2	3268.5	95.7	631	1 AAP93009	Aap93009 Genetic c
3	3268.5	95.7	631	1 AAP93009	Aap93009 Genetic c
4	3268.5	95.7	631	1 AAP93009	Aap93009 Genetic c
5	3219.5	94.3	729	1 AAP93008	Aap93008 Genetic c
6	3219.5	94.3	729	1 AAP93008	Aap93008 Genetic c
7	3219.5	94.3	729	1 AAP93008	Aap93008 Genetic c
8	3209.5	94.0	729	1 AAP93008	Aap93008 Genetic c
9	3169	92.8	616	3 AAY51078	Aay51078 Human fus
10	3169	92.8	616	3 AAY51078	Aay51078 Human fus
11	3169	92.8	616	3 AAY51078	Aay51078 Human fus
12	3077	90.1	616	1 AAP93012	Aap93012 Genetic c
13	2268	66.4	534	2 AAB19511	Aab19511 CD4-IgG1
14	2189	64.1	435	2 AAR26530	Aar26530 Sequence
15	2156.5	63.12	450	6 AAB37574	Aab37574 Human ftd
16	2151	63.0	530	2 AAR26763	Aar26763 Human ftd
17	2151	63.0	530	2 AAR26763	Aar26763 Human ftd
18	2151	63.0	530	2 AAR26763	Aar26763 Human ftd
19	2151	63.0	530	2 AAR26763	Aar26763 Human ftd
20	2151	63.0	530	2 AAR26763	Aar26763 Human ftd
21	2151	63.0	530	2 AAR26763	Aar26763 Human ftd
22	2151	63.0	530	2 AAR26763	Aar26763 Human ftd
23	2077	60.8	432	2 AAR46578	Aar46578 Human/CD4
24	2077	60.8	432	2 AAR46578	Aar46578 Human/CD4
25	2077	60.8	432	2 AAR46578	Aar46578 Human/CD4

26	2077	60.8	432	4 AAB67322	Aab67322 CD4-gamma
27	2077	60.8	432	4 AAB67322	Aab67322 CD4-gamma
28	2077	60.8	432	4 AAB67322	Aab67322 CD4-gamma
29	2066	60.5	481	3 AAB19510	Aab19510 CD4-IgM f
30	2066	60.5	481	3 AAB19510	Aab19510 CD4-IgM f
31	2060	60.3	481	3 AAY51081	Aay51081 Human fus
32	2058	60.3	481	3 AAY51081	Aay51081 Human fus
33	2053.5	60.1	436	1 AAP93011	Aap93011 Genetic c
34	2047.5	60.0	474	3 AAY51080	Aay51080 Human fus
35	2039	59.7	532	2 AAR27278	Aar27278 CD4-gamma
36	2039	59.7	532	2 AAR27278	Aar27278 CD4-gamma
37	2039	59.7	532	2 AAR27278	Aar27278 CD4-gamma
38	2039	59.7	532	2 AAR27278	Aar27278 CD4-gamma
39	2039	59.7	532	2 AAR27278	Aar27278 CD4-gamma
40	2039	59.7	532	2 AAR27278	Aar27278 CD4-gamma
41	2039	59.7	532	2 AAR27278	Aar27278 CD4-gamma
42	2039	59.7	532	2 AAR27278	Aar27278 CD4-gamma
43	2039	59.7	532	2 AAR27278	Aar27278 CD4-gamma
44	2038	59.7	458	2 AAB07769	Aab07769 DNA encod
45	2035	59.6	462	2 AAR27277	Aar27277 CD4-eta p
46	2035	59.6	462	2 AAR27277	Aar27277 CD4-eta p
47	2035	59.6	462	2 AAR27277	Aar27277 CD4-eta p
48	2035	59.6	462	2 AAR27277	Aar27277 CD4-eta p
49	2035	59.6	462	2 AAR27277	Aar27277 CD4-eta p
50	2035	59.6	462	2 AAR27277	Aar27277 CD4-eta p
51	2030	59.5	458	1 AAP81990	Aap81990 Clone pT4
52	2030	59.5	458	1 AAP81990	Aap81990 Clone pT4
53	2030	59.5	458	1 AAP81990	Aap81990 Clone pT4
54	2030	59.5	458	1 AAP81990	Aap81990 Clone pT4
55	2030	59.5	458	1 AAP81990	Aap81990 Clone pT4
56	2030	59.5	458	1 AAP81990	Aap81990 Clone pT4
57	2029	59.4	398	2 AAR78673	Aar78673 CD4 domain
58	2026	59.3	398	2 AAR78673	Aar78673 CD4 domain
59	2024	59.3	416	3 AAB19509	Aab19509 CD4-IgM f
60	2024	59.3	458	3 AAY88329	Aay88329 T4 glycop
61	2024	59.3	458	3 AAY88329	Aay88329 T4 glycop
62	2024	59.3	458	3 AAY88329	Aay88329 T4 glycop
63	2024	59.3	458	3 AAY88329	Aay88329 T4 glycop
64	2024	59.3	458	3 AAY88329	Aay88329 T4 glycop
65	2024	59.3	458	3 AAY88329	Aay88329 T4 glycop
66	2022	59.2	458	2 AAR13491	Aar13491 Human CD4
67	2018	59.1	394	1 AAP93506	Aap93506 Derived s
68	2017	59.1	402	1 AAP91922	Aap91922 Sequence
69	2017	59.1	402	1 AAP91922	Aap91922 Sequence
70	2016	59.1	458	4 AAG79087	Aag79087 Amino aci
71	2015	59.0	394	3 AAY88328	Aay88328 T4 glycop
72	2012	58.9	394	2 AAY39825	Aay39825 Soluble h
73	2006	58.8	458	7 ADE65841	Ad65841 Human CD4
74	2002.5	58.7	2458	2 AAR04031	Aar04031 Full leng
75	2002.5	58.7	2458	2 AAR04031	Aar04031 Full leng
76	2002	58.6	458	2 AAR06373	Aar06373 T4 encode
77	2002	58.6	524	1 AAP94703	Aap94703 Sequence
78	2001	58.6	394	2 AAB07768	Aab07768 The solub
79	2000	58.6	399	2 AAR20151	Aar20151 Chimpane
80	2000	58.6	458	2 AAR11285	Aar11285 gp120 bin
81	2000	58.6	458	2 AAR10988	Aar10988 Chimpane
82	1995	58.4	400	2 AAR20150	Aar20150 Chimpane
83	1994	58.4	458	2 AAR04910	Aar04910 T4 protei
84	1982	58.1	399	1 AAP93010	Aap93010 Genetic c
85	1955	57.3	400	2 AAR20154	Aar20154 Sol. rhes
86	1930	56.5	458	2 AAR15149	Aar15149 CD4 coord
87	1928.5	56.5	729	2 AAR41042	Aar41042 CD4-GBPR
88	1917	56.2	384	1 AAP90833	Aap90833 Amino aci
89	1915	56.1	942	1 AAR41041	Aar41041 CD4-GBPR
90	1911	56.0	1786	1 AAR41043	Aar41043 CD4-GBPR
91	1904	55.8	434	1 AAP96151	Aap96151 Sequence
92	1904	55.8	434	1 AAP96151	Aap96151 Sequence
93	1903	55.7	375	2 AAR07721	Aar07721 Recombina
94	1901	55.7	433	2 AAM41376	Aam41376 Human CD4
95	1896	55.5	433	2 AAY54500	Aay54500 Amino aci
96	1893	55.4	432	2 AAR74222	Aar74222 Epticope o
97	1893	55.4	432	2 AAY30514	Aay30514 Predicted
98	1893	55.4	432	2 ADA25188	Ada25188 CD4 eptico

99	1891	55.4	369	2	AAV39824	AAV39824 Soluble h
100	1891	55.4	369	3	AAV88327	AAV88327 T4 glycop
101	1891	55.4	370	1	AAV93528	AAV93528 Human sol
102	1891	55.4	370	4	AAV83356	AAV83356 Human CD4
103	1879	55.0	370	2	AAV12956	AAV12956 Lys(64) C
104	1878	55.0	370	2	AAV12954	AAV12954 His(40) C
105	1878	55.0	370	2	AAV12964	AAV12964 Val(143)
106	1878	55.0	370	2	AAV12965	AAV12965 Arg(150)
107	1877	55.0	370	2	AAV12958	AAV12958 Glu(80) C
108	1874	54.9	370	2	AAV12951	AAV12951 Glu(17)
109	1872	54.8	370	2	AAV12957	AAV12957 Asn(72)
110	1871	54.8	370	2	AAV12959	AAV12959 Asn(88)
111	1871	54.8	370	2	AAV12962	AAV12962 Leu(127)
112	1870	54.8	435	1	AAV90992	AAV90992 Human CD4
113	1866	54.7	370	2	AAV12966	AAV12966 Thr(162)
114	1865	54.6	370	2	AAV12952	AAV12952 Ile(23)
115	1864	54.6	370	2	AAV12967	AAV12967 Asp(155)
116	1863	54.6	435	1	AAV91289	AAV91289 Sequence
117	1862	54.5	370	2	AAV12955	AAV12955 Gly(48)
118	1861	54.5	370	2	AAV12963	AAV12963 His(132)
119	1861	54.5	370	2	AAV12960	AAV12960 Lys(99)
120	1859	54.5	433	2	AAV08335	AAV08335 CD4_3/20
121	1858	54.4	370	2	AAV12961	AAV12961 Ser(121)
122	1858	54.4	400	2	AAV20149	AAV20149 Sol. rh
123	1857	54.4	458	2	AAV10987	AAV10987 Rheus mo
124	1856	54.4	458	2	AAV20148	AAV20148 Cynomolgu
125	1852	54.2	370	2	AAV12953	AAV12953 Thr(27)

ALIGNMENTS

RESULT 1

AAP93009 standard; protein: 631 AA.

ID	AAP93009	standard; protein: 631 AA.
AC	AAP93009;	
XX	25-MAR-2003	(revised)
DT	02-NOV-1992	(first entry)
DE	Genetic construct which encodes CD4 linked to human IgG1 at the Esp site	
DE	upstream of the hinge region (fusion protein CD4E-gamma-1).	
XX		
KW	Fusion protein: immunoglobulin-like molecule: HIV; SIV; therapy:	
XX	diagnosis; CD4; gp120; binding fragment; glycoprotein; variable region.	
XX		
OS	Homo sapiens.	
XX		
PN	EP325262-A.	
XX		
PD	26-JUL-1989.	
XX		
PF	20-JAN-1989;	89EP-00100913.
XX		
PR	22-JAN-1988;	88US-00147351.
XX		
PA	(GENO) GEN HOSPITAL CORP.	
XX		
PI	Seed B;	
XX		
DR	WPI; 1989-214472/30.	
DR	N-PSDB; AAN90357.	
XX		
PT	Immunoglobulin-CD4 fusion proteins - used for treating HIV or SIV	
PT	infections or detecting HIV or SIV in sample.	
XX		
PS	Example; Table 2, Page 24-33; 68pp; English.	
XX		
CC	The fusion protein genes of the invention pref. comprise cDNA sequences	
CC	which encode CD4 or a fragment which binds gp120 ligated to an expression	
CC	plasmid which encodes an antibody in which the variable region of the	
CC	gene has been deleted (see WO87-02671). The CD4 portion of the fusion	

CC protein may comprise the complete CD4 sequence, the 370 AA extracellular
 CC region and the membrane spanning domain, or the extracellular region. The
 CC Ig heavy chain is pref. from IgM, IgG1 or IgG3. The following are
 CC specifically claimed: fusion proteins CD4E-gamma-1, CD4Mmu, CD4Pmu, CD4E-
 CC gamma1, and CD4Mmu (No. 67608), pCD4P-gamma (No. 67609) and pCD4E-gamma-1
 CC (No. 67610). The plasmid containing (pCD4E-gamma-1) has been deposited in
 CC E. coli (MC1061/P3) at the ATCC under accession number 67610. (Updated on
 CC 25-MAR-2003 to correct PA field.)

SO Sequence 631 AA;

Query Match 95.7%; Score 3268.5; DB 1; Length 631;
 Best Local Similarity 99.0%; Pred. No. 1.8e-170;
 Matches 624; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

QY	1	MNRGVPFRHLIVLQALPAATQGNVVLGKGDVETLCTASOKKSIOFHMKNNOIK	60
DB	1	MNRGVPFRHLIVLQALPAATQGNVVLGKGDVETLCTASOKKSIOFHMKNNOIK	60
QY	61	ILNGQSFLLTKGSPKLNDRADSRSLMDQGNFPLIINKLTIEDSDTYICEVEDQKEVQL	120
DB	61	ILNGQSFLLTKGSPKLNDRADSRSLMDQGNFPLIINKLTIEDSDTYICEVEDQKEVQL	120
QY	121	LVFGLTANSDTHLLQGGSLTLLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180
DB	121	LVFGLTANSDTHLLQGGSLTLLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180
QY	181	TWTCVTIQQNKVEFKDIIVLAFQKASSTVYKKEGQVSPFLATFVELTSSGGLMW	240
DB	181	TWTCVTIQQNKVEFKDIIVLAFQKASSTVYKKEGQVSPFLATFVELTSSGGLMW	240
QY	241	QABRASSSKSMITFDLKNKEVSRYVQDPKLCQKGLPLHLTPQALPOYAGSGNTLA	300
DB	241	QABRASSSKSMITFDLKNKEVSRYVQDPKLCQKGLPLHLTPQALPOYAGSGNTLA	300
QY	301	LEAKTGKHOEVLVVRATQLOKNTLCBYWGPTSPKMLSLKLENKAKVSKREKVMV	360
DB	301	LEAKTGKHOEVLVVRATQLOKNTLCBYWGPTSPKMLSLKLENKAKVSKREKVMV	360
QY	361	LNPEAGWQCLLSGQVLLSNTKVLPTWSTPVHADPEBPKSCDKHTCPCPAPBLG	420
DB	361	LNPEAGWQCLLSGQVLLSNTKVLPTWSTPVHADPEBPKSCDKHTCPCPAPBLG	420
QY	416	GPSVFLFPPKPKOTLMSTRPEVTCVVVDVSHEDPEVKFMVYDGVENHNAKTRPREQY	475
DB	421	GPSVFLFPPKPKOTLMSTRPEVTCVVVDVSHEDPEVKFMVYDGVENHNAKTRPREQY	480
QY	476	NSTYRVVSVLTVLHQMVLNGEKYCKVSNKALPAPIEKTIISKAKGPREPOVYTLPPSRD	535
DB	481	NSTYRVVSVLTVLHQMVLNGEKYCKVSNKALPAPIEKTIISKAKGPREPOVYTLPPSRD	540
QY	536	ELTKNOVSLTCLVKGFPSPDIAVWESNGQPENNYKTPPTLSDGSPFLYSKLTIVKSR	595
DB	541	ELTKNOVSLTCLVKGFPSPDIAVWESNGQPENNYKTPPTLSDGSPFLYSKLTIVKSR	600
QY	596	WQGNVFSQSVMEALNHYTKSLSLSPG	625
DB	601	WQGNVFSQSVMEALNHYTKSLSLSPG	630

RESULT 2

AAB19508 standard; protein: 631 AA.

ID	AAB19508	standard; protein: 631 AA.
AC	AAB19508;	
XX	09-JAN-2001	(first entry)
DT		
DE	CD4-IgG1 fusion protein CH4Bgammal.	
XX		
KW	CD4; IgG1; human; CD4Bgammal; fusion protein; immunoglobulin; HIV; SIV;	
KW	gp120; therapy; diagnosis.	
XX		


```

OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT Protein 1..395 "CD4 extracellular region"
XX FT Protein /note="CD4 extracellular region"
XX FT Protein 400..631
XX FT Protein /note="IgG1 heavy chain"
XX
XX US6117656-A.
XX
XX 12-SEP-2000.
XX
XX 07-JUN-1995; 95US-00479353.
XX
XX 22-JAN-1988; 88US-00147351.
XX 23-JAN-1989; 89US-00299596.
XX 09-JUN-1992; 92US-00896781.
XX 12-APR-1993; 93US-00057952.
XX 04-FEB-1994; 94US-00191708.
XX
XX (GEHO ) GEN HOSPITAL CORP.
XX
XX Seed B;
XX
XX WPI; 2000-586558/55.
XX DR N-PSDB; AAA50661.
XX
XX CD4-immunoglobulin fusion proteins, useful for targeting gp120 of HIV or
XX SIV.
XX
XX Example 1; Col 29-42; 39pp; English.
XX
XX The present sequence is that of fusion protein CD4Egammal comprising the
XX extracellular portion of CD4, which binds to HIV gp120, linked at its C-
XX terminus to the human IgG1 heavy chain. To obtain the fusion protein, DNA
XX encoding CD4 was linked to IgG1 DNA at the Esp site upstream of the hinge
XX region (see AAA50661). Fusion protein CD4Egammal and a nucleic acid
XX encoding it are claimed. Also claimed are a vector comprising the nucleic
XX acid, and a method of producing the fusion protein in secreted form using
XX a transformed host cell. The fusion protein may further comprise a
XX therapeutic agent, radiolabel or NMR imaging agent. The fusion protein
XX can be administered to an animal (including humans) for treatment of HIV
XX or SIV infection, and can also be used in assays for HIV or SIV, imaging
XX and tissue stains. IgG1 fusion proteins such as CD4Egammal provide both
XX complement-mediated and cell-mediated immunity
XX
XX Sequence 631 AA;
XX
XX Query Match 95.7%; Score 3268.5; DB 3; Length 631;
XX Best Local Similarity 99.0%; Pred. No. 1.8e-170;
XX Matches 624; Conservative 0; Mismatches 1; Indels 5; Gaps 1;
XX
XX 1 MNRGVFRHLILVLAQALPAATQGNKVVYLGKKGDTVELTCTASQKKSIOFMKNSNQIK 60
XX 1 MNRGVFRHLILVLAQALPAATQGNKVVYLGKKGDTVELTCTASQKKSIOFMKNSNQIK 60
XX
XX 61 ILGNQSSFLTKGSKNDRAADSRSLMDQGNFPLIKNIKIESDPTICEVEDQKEEVL 120
XX 61 ILGNQSSFLTKGSKNDRAADSRSLMDQGNFPLIKNIKIESDPTICEVEDQKEEVL 120
XX
XX 121 LVFGLTANSDTHLLOQOSLTLTLESPPGSSPVQCSPRGKNIQSGKTLVSQLELDQSG 180
XX 121 LVFGLTANSDTHLLOQOSLTLTLESPPGSSPVQCSPRGKNIQSGKTLVSQLELDQSG 180
XX
XX 181 TWCTVLAQKQKVEFKIDIVLAFOKASSIVYKKEGEQVFEFPLAFTVEKLTGSGELMW 240
XX 181 TWCTVLAQKQKVEFKIDIVLAFOKASSIVYKKEGEQVFEFPLAFTVEKLTGSGELMW 240
XX
XX 241 QMRASSSSKSWITFDLKKEVSVKRTOPKLOMGKKLPLHLTLPALQYVAGSGNLTIA 300
XX 241 QMRASSSSKSWITFDLKKEVSVKRTOPKLOMGKKLPLHLTLPALQYVAGSGNLTIA 300
XX
XX 301 LEAKTGKLHQEVNLVVMRATOLQKNLTCEVMGPTSPKMLSLKLEKKAIVSKREKPVWV 360

```

```

DB 301 LEAKTGKLHQEVNLVVMRATOLQKNLTCEVMGPTSPKMLSLKLEKKAIVSKREKPVWV 360
XX
XX 361 LMPKAGMOCCLSDSGOVLBSNIVLPTWSTPV-----EPKSCKTKTTPCPAPBELLG 415
XX 361 LMPKAGMOCCLSDSGOVLBSNIVLPTWSTPVHADDEEPKSCDKTKTTPCPAPBELLG 420
XX
XX 416 GPSVFLPPPKDPTLMISRTPEVTCVVVDVSHEDPEVKFNMYVDQVFNAAKTKEDEQY 475
XX 421 GPSVFLPPPKDPTLMISRTPEVTCVVVDVSHEDPEVKFNMYVDQVFNAAKTKEDEQY 480
XX
XX 476 NSTYRVSVLTIVLHODMNGKEKCKVSNKALPAPIEKTISKAKGQPREPPVYTLPSPSD 535
XX 481 NSTYRVSVLTIVLHODMNGKEKCKVSNKALPAPIEKTISKAKGQPREPPVYTLPSPSD 540
XX
XX 536 ELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTTPVLDSGSEFYLKLTVDKSR 595
XX 541 ELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTTPVLDSGSEFYLKLTVDKSR 600
XX
XX 596 WQGGNVFSCSVMHKALHNHYTQKSLSLSPG 625
XX 601 WQGGNVFSCSVMHKALHNHYTQKSLSLSPG 630
XX
XX RESULT 3
XX AA51079
XX ID AA51079 standard; protein; 631 AA.
XX
XX AA51079;
XX
XX 23-MAR-2000 (first entry)
XX
XX Human fusion protein CD4Egammal.
XX
XX Fusion protein; human; CD4; IgG1; immunoglobulin; gp120;
XX anti-human immunodeficiency virus; CD4Egammal.
XX
XX OS Homo sapiens.
XX Synthetic.
XX
XX US6004781-A.
XX
XX 21-DEC-1999.
XX
XX 04-FEB-1994; 94US-00191708.
XX
XX 22-JAN-1988; 88US-00147351.
XX 23-JAN-1989; 89US-00299596.
XX 09-JUN-1992; 92US-00896781.
XX 12-APR-1993; 93US-00057952.
XX
XX (GEHO ) GEN HOSPITAL CORP.
XX
XX Seed B;
XX
XX WPI; 2000-085792/07.
XX DR N-PSDB; AA244062.
XX
XX Fusion protein useful for the treatment of human immunodeficiency virus.
XX
XX Example 1; Col 29-42; 39pp; English.
XX
XX This invention describes a novel nucleic acid (I) encoding a fusion
XX protein comprising a DNA sequence encoding amino acids 1-173 of CD4 (II)
XX and a DNA sequence encoding a human immunoglobulin (Ig) heavy or light
XX chain (III). The products of the invention have anti-human
XX immunodeficiency virus (HIV) activity and are capable of binding to
XX gp120. The fusion protein is useful for treating human immunodeficiency
XX virus (HIV) or simian immunodeficiency virus (SIV). This sequence
XX represents the fusion protein CD4Egammal which is constructed from CD4
XX linked to human IgG1 upstream of the hinge region
XX
XX Sequence 631 AA;

```

Query Match 95.7%; Score 3268.5; DB 3; Length 631;
 Best Local Similarity 99.0%; Pred. No. 1.8e-170;
 Matches 624; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

```

QY 1 MNRGVPFRHLLVLTQALLPAATQGNKVLGKGGDTVELCTASQKSIQPHMKNNOIK 60
DB 1 MNRGVPFRHLLVLTQALLPAATQGNKVLGKGGDTVELCTASQKSIQPHMKNNOIK 60
QY 61 ILNGQSFLLTKGPSKLNDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILNGQSFLLTKGPSKLNDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDTHLLOGSLLTLTLESPGSSPSVQCRSPRKNIOGKTLISVSOLELDOSG 180
DB 121 LVFGLTANSDTHLLOGSLLTLTLESPGSSPSVQCRSPRKNIOGKTLISVSOLELDOSG 180
QY 181 TWCTVLONQKKVEFKIDIVLAFQKASSIYKKGEQVFSPLAFTVEKLTGSGELMW 240
DB 181 TWCTVLONQKKVEFKIDIVLAFQKASSIYKKGEQVFSPLAFTVEKLTGSGELMW 240
QY 241 QAERASSKSMITFDLKNKEVSVKRVTDPKLQWKKLPLHLTLPOALPOYAGSGLTLA 300
DB 241 QAERASSKSMITFDLKNKEVSVKRVTDPKLQWKKLPLHLTLPOALPOYAGSGLTLA 300
QY 301 LEAKTGKLEHGVNLVVMRATOLQKNTLCEVWGPTSPKMLSLKLENKEAVSKREKPYW 360
DB 301 LEAKTGKLEHGVNLVVMRATOLQKNTLCEVWGPTSPKMLSLKLENKEAVSKREKPYW 360
QY 361 LNPBAGMMOCLLSDSGVLLSESNIKVLPWTSTPV-----EPKSCDKHTGCPCPAPELLG 415
DB 361 LNPBAGMMOCLLSDSGVLLSESNIKVLPWTSTPVADDEEPKSCDKHTGCPCPAPELLG 420
QY 416 GPSVFLPPPKPDKTLMISRTPEVTCVVVDVSHEDPEVKFNNYVGDVEVHNAKTKRPREQY 475
DB 416 GPSVFLPPPKPDKTLMISRTPEVTCVVVDVSHEDPEVKFNNYVGDVEVHNAKTKRPREQY 480
QY 476 NSTYRVSVLTITVLDMLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRD 535
DB 476 NSTYRVSVLTITVLDMLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRD 540
QY 536 ELTKNQVSLTCLVKGFYPSDIAVEMESNGQPENNYKTTTPVLDSGSEFLLYSKLTVDKSR 595
DB 536 ELTKNQVSLTCLVKGFYPSDIAVEMESNGQPENNYKTTTPVLDSGSEFLLYSKLTVDKSR 600
QY 596 WOQGNVFSCSVMHEALHNHYTQKSLSLSPG 625
DB 601 WOQGNVFSCSVMHEALHNHYTQKSLSLSPG 630

```

RESULT 4
 AA59169
 ID AA59169 standard; protein; 631 AA.
 AC AA59169;
 DE 14-MAR-2000 (first entry)
 DT CD4-Ig fusion protein CD4Bgamma1.
 KW HIV; extracellular; CD4; gp120; immunoglobulin; Ig; fusion protein;
 XX secreted protein; HIV infection; medicament.
 OS Synthetic.
 OS Homo sapiens.
 PN CA1340741-C.
 PD 14-SEP-1999.
 PF 20-JAN-1989; 89CA-00588749.
 XX 20-JAN-1989; 89CA-00588749.

XX (GEO) GEN HOSPITAL CORP.
 PA Seed B;
 XX WPI; 2000-063015/06.
 DR N-PSDB; AA248202.
 XX New fusion gene encoding immunoglobulin-CD4 fusion proteins, useful in
 PT the treatment of HIV or simian immunodeficiency virus infections.
 PS Example 1; Page 37-46; 89pp; English.
 XX The invention provides a fusion gene encoding a fusion protein that
 CC comprises an extracellular CD4 DNA sequence or its fragment which binds
 CC to HIV gp120 when fused to an immunoglobulin (Ig) chain and the DNA
 CC sequence of an Ig heavy or light chain, where the DNA sequence encoding
 CC the variable region has been replaced with the DNA sequence which encodes
 CC extracellular CD4 or its gp120 binding fragment. The fusion protein is
 CC capable of being secreted. The fusion proteins are useful for treating
 CC HIV or HIV infections in animals, preferably humans. They are also useful
 CC for producing medicaments which can be used for creating HIV or HIV
 CC infections in humans. The present sequence represents the fusion protein
 CC CD4Bgamma1 where the CD4 is linked to human IgG1 at the Esp site upstream
 CC of the hinge region
 XX Sequence 631 AA:

Query Match 95.7%; Score 3268.5; DB 3; Length 631;
 Best Local Similarity 99.0%; Pred. No. 1.8e-170;
 Matches 624; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

```

QY 1 MNRGVPFRHLLVLTQALLPAATQGNKVLGKGGDTVELCTASQKSIQPHMKNNOIK 60
DB 1 MNRGVPFRHLLVLTQALLPAATQGNKVLGKGGDTVELCTASQKSIQPHMKNNOIK 60
QY 61 ILNGQSFLLTKGPSKLNDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILNGQSFLLTKGPSKLNDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDTHLLOGSLLTLTLESPGSSPSVQCRSPRKNIOGKTLISVSOLELDOSG 180
DB 121 LVFGLTANSDTHLLOGSLLTLTLESPGSSPSVQCRSPRKNIOGKTLISVSOLELDOSG 180
QY 181 TWCTVLONQKKVEFKIDIVLAFQKASSIYKKGEQVFSPLAFTVEKLTGSGELMW 240
DB 181 TWCTVLONQKKVEFKIDIVLAFQKASSIYKKGEQVFSPLAFTVEKLTGSGELMW 240
QY 241 QAERASSKSMITFDLKNKEVSVKRVTDPKLQWKKLPLHLTLPOALPOYAGSGLTLA 300
DB 241 QAERASSKSMITFDLKNKEVSVKRVTDPKLQWKKLPLHLTLPOALPOYAGSGLTLA 300
QY 301 LEAKTGKLEHGVNLVVMRATOLQKNTLCEVWGPTSPKMLSLKLENKEAVSKREKPYW 360
DB 301 LEAKTGKLEHGVNLVVMRATOLQKNTLCEVWGPTSPKMLSLKLENKEAVSKREKPYW 360
QY 361 LNPBAGMMOCLLSDSGVLLSESNIKVLPWTSTPV-----EPKSCDKHTGCPCPAPELLG 415
DB 361 LNPBAGMMOCLLSDSGVLLSESNIKVLPWTSTPVADDEEPKSCDKHTGCPCPAPELLG 420
QY 416 GPSVFLPPPKPDKTLMISRTPEVTCVVVDVSHEDPEVKFNNYVGDVEVHNAKTKRPREQY 475
DB 416 GPSVFLPPPKPDKTLMISRTPEVTCVVVDVSHEDPEVKFNNYVGDVEVHNAKTKRPREQY 480
QY 476 NSTYRVSVLTITVLDMLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRD 535
DB 476 NSTYRVSVLTITVLDMLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRD 540
QY 536 ELTKNQVSLTCLVKGFYPSDIAVEMESNGQPENNYKTTTPVLDSGSEFLLYSKLTVDKSR 595
DB 536 ELTKNQVSLTCLVKGFYPSDIAVEMESNGQPENNYKTTTPVLDSGSEFLLYSKLTVDKSR 600
QY 596 WOQGNVFSCSVMHEALHNHYTQKSLSLSPG 625

```

```

Db      601 MOGNVFCSCVMHEALHNHYTQKSLSPG 630
|||||
RESULT 5
ID      AAP93008 standard; protein; 729 AA.
XX
XX      AAP93008;
AC
XX      25-MAR-2003 (revised)
XX      02-NOV-1992 (first entry)
DT
DE      Genetic construct which encodes CD4 linked to human IgG1 at the Hind3
XX      site upstream of the CH1 region (fusion protein CD4H-gamma-1).
XX
XX      Fusion protein; immunoglobulin-like molecule; HIV; SIV; therapy;
XX      diagnosis; CD4; gp120; binding fragment; glycoprotein; variable region.
OS      Homo sapiens.
XX
XX      EP325262-A.
XX
XX      26-JUL-1989.
XX
XX      20-JAN-1989; 89EP-00100913.
XX
XX      22-JAN-1988; 88US-00147351.
XX
XX      (GEHO ) GEN HOSPITAL CORP.
PA
PI      Seed B;
XX
XX      WPI: 1989-214472/30.
XX      N-PSDB; AAN90356.
DR
PT      Immunoglobulin-CD4 fusion proteins - used for treating HIV or SIV
XX      infections or detecting HIV or SIV in sample.
XX
XX      Example; Table 1, Page 12-23; 68pp; English.
PS
XX
XX      The fusion protein genes of the invention pref. comprise cDNA sequences
XX      CC which encode CD4 or a fragment which binds gp120 ligated to an expression
XX      CC plasmid which encodes an antibody in which the variable region of the
XX      CC gene has been deleted (see WO87-02671). The CD4 portion of the fusion
XX      CC protein may comprise the complete CD4 sequence, the 370 AA extracellular
XX      CC region and the membrane spanning domain, or the extracellular region. The
XX      CC Ig heavy chain is pref. from IgM, IgG1 or IgG3. The following are
XX      CC specifically claimed: fusion proteins CD4H-gamma-1, CD4Mmu, CD4Pmu, CD4E-
XX      CC gamma1, and CD4Mmu (No.67608), pCD4P-gamma (No.67609) and pCD4E-gamma-1
XX      CC (No.67610). The plasmid containing (pCD4H-gamma-1) has been deposited in
XX      CC E. coli (MC1061/P3) at the ATCC under accession number 67611. (Updated on
XX      CC 25-MAR-2003 to correct PA field.)
XX
XX      Sequence 729 AA;
SQ
Query Match      94.3%; Score 3219.5; DB 1; Length 729;
Best Local Similarity 85.7%; Pred. No.1e-167;
Matches 624; Conservative 0; Mismatches 1; Indels 103; Gaps 1;

```

```

Db      181 TWCTCTVLTQKQKVEFKIDIVLAFQKASSIYVKKEGEQVRSFPLAFVTEKLTGSQELMW 240
|||||
Db      241 QAERASSSSKSWITPDLNKRKEVSVKRVYTDPPKLGKPKLPLHLTLPOALPOYAGSNLTILA 300
|||
Db      241 QAERASSSSKSWITPDLNKRKEVSVKRVYTDPPKLGKPKLPLHLTLPOALPOYAGSNLTILA 300
|||
Db      301 LEAKTKGLHGVNLVVMRATOLQKNTLCEVNGPSTPKMLSLKENKEAKYSKREKPYWV 360
|||
Db      301 LEAKTKGLHGVNLVVMRATOLQKNTLCEVNGPSTPKMLSLKENKEAKYSKREKPYWV 360
|||
Db      361 LNPEAGMWCQLLSDSGQVLTLESNIKVLPTWSTP----- 393
|||
Db      361 LNPEAGMWCQLLSDSGQVLTLESNIKVLPTWSTPYNADPEASTKGPSVPLAPSSKSTSGG 420
|||
Db      394 ----- 393
|||
Db      421 TAAIGCLVKDYFPEPEVTVMNSGALLTSGVHTPPPAVLQSSGLYSSTVTPVSSSLGTQTY 480
|||
Db      394 -----VEPKSCDKHTHTCPCPAPRLTGGPSVFLPFPKPKDTLMTSRTPE 437
|||
Db      481 ICNVNHRKSNTRKVDKRVKPKSCDKHTHTCPCPAPRLTGGPSVFLPFPKPKDTLMTSRTPE 540
|||
Db      438 VTCVVVDVSHEDPEVKFMWYDGVVEVNAKTRPREQYNSTYRVVSVLTVLHQDWLNGKE 497
|||
Db      541 VTCVVVDVSHEDPEVKFMWYDGVVEVNAKTRPREQYNSTYRVVSVLTVLHQDWLNGKE 600
|||
Db      498 YKCKVSNKALPAPIEKITISKAKGPREPOVYTLTPSRDELTKNQVSLTCLYKGFPSDIA 557
|||
Db      601 YKCKVSNKALPAPIEKITISKAKGPREPOVYTLTPSRDELTKNQVSLTCLYKGFPSDIA 660
|||
Db      558 VEMESNGQEPNNYKTPPVLDSDGSFFLYSKLTVDKSRWQQGNVSCVMHEALHNHYTQ 617
|||
Db      661 VEMESNGQEPNNYKTPPVLDSDGSFFLYSKLTVDKSRWQQGNVSCVMHEALHNHYTQ 720
|||
Db      618 KSLSLSPG 625
|||
Db      721 KSLSLSPG 728
|||
RESULT 6
AAB19507
ID      AAB19507 standard; protein; 729 AA.
XX
XX      AAB19507;
AC
XX      09-JAN-2001 (first entry)
XX
XX      CD4-IgG1 fusion protein CH4Hgamma1.
DE
XX      CD4-IgG1 fusion protein CH4Hgamma1.
XX
XX      CD4; IgG1; human; CD4Hgamma1; fusion protein; immunoglobulin; HIV; SIV;
XX      gp120; therapy; diagnosis.
XX
XX      Homo sapiens.
OS
XX
XX      Key
XX      FH      Location/Qualifiers
XX      FT      Protein
XX      FT      /note="CD4 extracellular region"
XX      FT      Protein
XX      FT      /note="IgG1 heavy chain"
XX
XX      US6117656-A.
XX
XX      12-SEP-2000.
XX
XX      07-JUN-1995; 95US-00479353.
XX
XX      22-JAN-1988; 88US-00147351.
XX      23-JAN-1989; 89US-00299596.
XX      09-JUN-1992; 92US-00896781.
XX      12-APR-1993; 93US-00057952.
XX      04-FEB-1994; 94US-00191708.
XX

```

PA (GEMO) GEN HOSPITAL CORP.
 XX Seed B;
 PI WPI: 2000-586558/55.
 DR N-PSDB; AAA50660.
 XX CD4-immunoglobulin fusion proteins, useful for targeting gp120 of HIV or
 PT SIV.
 XX Example 1; Col 13-30; 39pp; English.
 PS
 XX The present sequence is that of fusion protein CD4Hgmam1 comprising the
 CC extracellular portion of CD4, which binds to HIV gp120, linked at its C-
 CC terminus to the human IgG1 heavy chain. To obtain the fusion protein, DNA
 CC encoding CD4 was linked to IgG1 DNA at the Hind3 site upstream of the CH1
 CC region (see AAA50660). Fusion protein CD4Hgmam1 and a nucleic acid
 CC encoding it are claimed. Also claimed are a vector comprising the nucleic
 CC acid, and a method of producing the fusion protein in secreted form using
 CC a transformed host cell. The fusion protein may further comprise a
 CC therapeutic agent, radiolabel or NMR imaging agent. The fusion protein
 CC can be administered to an animal (including humans) for treatment of HIV
 CC or SIV infection, and can also be used in assays for HIV or SIV, imaging
 CC and tissue stains. IgG1 fusion proteins such as CD4Hgmam1 provide both
 CC complement-mediated and cell-mediated immunity
 XX
 SQ Sequence 729 AA;
 Query Match 94.3%; Score 3219.5; DB 3; Length 729;
 Best Local Similarity 85.7%; Pred. No. 1e-167;
 Matches 624; Conservative 0; Mismatches 1; Indels 103; Gaps 1;
 QY 1 MNRGVPFRHLIVLQALLPAATQGNKVVLGKGGDTVELTCTASQKSIQFHMKNNOIK 60
 DB 1 MNRGVPFRHLIVLQALLPAATQGNKVVLGKGGDTVELTCTASQKSIQFHMKNNOIK 60
 QY 61 ILNGSGFLTKGPKSLNDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEEVQL 120
 DB 61 ILNGSGFLTKGPKSLNDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEEVQL 120
 QY 121 LVFGITANSDTHLLOGSITLTLESPPGSSPSVQCRSPRGKNIQGGKTLSTLSQLELQDSG 180
 DB 121 LVFGITANSDTHLLOGSITLTLESPPGSSPSVQCRSPRGKNIQGGKTLSTLSQLELQDSG 180
 QY 121 LVFGITANSDTHLLOGSITLTLESPPGSSPSVQCRSPRGKNIQGGKTLSTLSQLELQDSG 180
 DB 121 LVFGITANSDTHLLOGSITLTLESPPGSSPSVQCRSPRGKNIQGGKTLSTLSQLELQDSG 180
 QY 181 TWTCTVLNOKKVEFKIDIVLAFQKASSIYKKEGBOVESFPFAFTVEKLTGSGELMW 240
 DB 181 TWTCTVLNOKKVEFKIDIVLAFQKASSIYKKEGBOVESFPFAFTVEKLTGSGELMW 240
 QY 181 TWTCTVLNOKKVEFKIDIVLAFQKASSIYKKEGBOVESFPFAFTVEKLTGSGELMW 240
 DB 181 TWTCTVLNOKKVEFKIDIVLAFQKASSIYKKEGBOVESFPFAFTVEKLTGSGELMW 240
 QY 241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKQLQMGKULPLHLTLPOALPOYAGSGLTLA 300
 DB 241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKQLQMGKULPLHLTLPOALPOYAGSGLTLA 300
 QY 301 LEAKTGKILHOEVNLVYMRATQQLKLTCEWGPSTPKMLSLKLNKAKYSKREKPVWV 360
 DB 301 LEAKTGKILHOEVNLVYMRATQQLKLTCEWGPSTPKMLSLKLNKAKYSKREKPVWV 360
 QY 361 LNPBAGMMQCLSDSGOVLLESGINIVLPTWSTP----- 393
 DB 361 LNPBAGMMQCLSDSGOVLLESGINIVLPTWSTP----- 393
 QY 394 ----- 393
 DB 394 ----- 393
 QY 421 TAAAGCLVSYRPEPEVTVSMNSGALTSGVTFPAVLQSSGLYSLSVTVTPSSSLGTQTY 480
 DB 421 TAAAGCLVSYRPEPEVTVSMNSGALTSGVTFPAVLQSSGLYSLSVTVTPSSSLGTQTY 480
 QY 434 -----VERPKSCDKTHTCTPCPAPBELLGSPSYFLFPPEPKDTLMSRTP 437
 DB 434 -----VERPKSCDKTHTCTPCPAPBELLGSPSYFLFPPEPKDTLMSRTP 437
 QY 481 ICNVNHHKSNTRKVDKKEVPEKSCDKTHTCTPCPAPBELLGSPSYFLFPPEPKDTLMSRTP 540
 DB 481 ICNVNHHKSNTRKVDKKEVPEKSCDKTHTCTPCPAPBELLGSPSYFLFPPEPKDTLMSRTP 540
 QY 497 VTCVVVDVSHEDPEVKFNNYVVDGVEVNAKTKPREEOYNSTRYVSVLTVLHODMLNKE 497
 DB 497 VTCVVVDVSHEDPEVKFNNYVVDGVEVNAKTKPREEOYNSTRYVSVLTVLHODMLNKE 497
 QY 541 VTCVVVDVSHEDPEVKFNNYVVDGVEVNAKTKPREEOYNSTRYVSVLTVLHODMLNKE 600
 DB 541 VTCVVVDVSHEDPEVKFNNYVVDGVEVNAKTKPREEOYNSTRYVSVLTVLHODMLNKE 600
 QY 498 YKCKVSNKALPAPIEKTISSAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIA 557

DB 601 YKCKVSNKALPAPIEKTISSAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIA 660
 QY 558 VEWESNGOPENNYKTPPVLDSDGSFLYSKLTVDKSRMQQGNVFSCSVMHEALHNHYTQ 617
 DB 661 VEWESNGOPENNYKTPPVLDSDGSFLYSKLTVDKSRMQQGNVFSCSVMHEALHNHYTQ 720
 QY 618 KSLSLSPG 625
 DB 721 KSLSLSPG 728
 RESULT 7
 AAY59168
 ID AAY59168 standard; protein; 729 AA.
 XX
 AC AAY59168;
 XX
 DT 14-MAR-2000 (first entry)
 XX
 DE CD4-Ig fusion protein CD4Hgmam1.
 XX
 KM HIV; extracellular; CD4; gp120; immunoglobulin; Ig; fusion protein;
 KM secreted protein; SIV infection; medicament.
 XX
 OS Synthetic.
 OS Homo sapiens.
 XX
 PM CA1340741-C.
 XX
 PD 14-SEP-1999.
 XX
 PF 20-JAN-1989; 89CA-00588749.
 XX
 PR 20-JAN-1989; 89CA-00588749.
 XX
 PA (GEMO) GEN HOSPITAL CORP.
 XX
 PI Seed B;
 XX
 DR WPI: 2000-063015/06.
 DR N-PSDB; AAZ48201.
 XX
 PT New fusion gene encoding immunoglobulin-CD4 fusion proteins, useful in
 PT the treatment of HIV or simian immunodeficiency virus infections.
 PS Example 1; Page 25-36; 89pp; English.
 XX
 XX The invention provides a fusion gene encoding a fusion protein that
 CC comprises an extracellular CD4 DNA sequence or its fragment which binds
 CC to HIV gp120 when fused to an immunoglobulin (Ig) chain and the DNA
 CC sequence of an Ig heavy or light chain, where the DNA sequence encoding
 CC the variable region has been replaced with the DNA sequence which encodes
 CC extracellular CD4 or its gp120 binding fragment. The fusion protein is
 CC capable of being secreted. The fusion proteins are useful for treating
 CC HIV or SIV infections in animals, preferably humans. They are also useful
 CC for producing medicaments which can be used for treating HIV or SIV
 CC infections in humans. The present sequence represents the fusion protein
 CC CD4Hgmam1 where the CD4 is linked to human IgG1 at the Hind3 site
 CC upstream of the CH1 region
 XX
 SQ Sequence 729 AA;
 Query Match 94.3%; Score 3219.5; DB 3; Length 729;
 Best Local Similarity 85.7%; Pred. No. 1e-167;
 Matches 624; Conservative 0; Mismatches 1; Indels 103; Gaps 1;
 QY 1 MNRGVPFRHLIVLQALLPAATQGNKVVLGKGGDTVELTCTASQKSIQFHMKNNOIK 60
 DB 1 MNRGVPFRHLIVLQALLPAATQGNKVVLGKGGDTVELTCTASQKSIQFHMKNNOIK 60
 QY 61 ILNGSGFLTKGPKSLNDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEEVQL 120